

Zdravotnictví v ČR

Technická zpráva k csv souborům dotazu C
2.části projektu do předmětu UPA

FIT VUT v Brně, 2021

Název týmu

Tým xloren15

Autoři

Jan Lorenc, Bc. (xloren15)

Marek Hlavačka, Bc. (xhlava50)

Martin Smetana, Ing. (xsmeta10)

Struktura neupraveného souboru

Z oborů zdravotní péče bylo zvoleno 20 nejčastějších dle aktuálního stavu. Data pro tyto obory se poté brala po čtvrtletích, tedy po třech měsících, počínaje aktuálním měsícem. Zajímavostí se ukázal být obor praktického lékařství, který se v listopadu 2019 přejmenoval z původního veřejného lékařství, čehož jsme si v první části projektu nevšimli, neboť to nikde není zaznamenáno. Data pro tento obor jsou tedy kombinací těchto byt' stejných, ale různě pojmenovaných oborů. Vyextrahovaná kolekce dat ve formátu csv vypadá následovně:

Obor péče	2018-11	2019-02	2019-05	2019-08	2019-11	2020-02	2020-05	2020-08	2020-11	2021-02	2021-05	2021-08	2021-11
Anesteziologie a intenzivní medicína	956	952	959	975	943	936	941	945	951	955	963	963	971
Chirurgie	1575	1567	1570	1574	1532	1498	1504	1502	1505	1505	1515	1522	1529
Dermatovenerologie	1069	1073	1076	1068	1056	1027	1031	1031	1030	1016	1016	1019	1026
Endokrinologie a diabetologie	994	994	998	1000	977	943	946	945	949	937	930	930	931
Fyzioterapeut	3045	3072	3115	3146	3177	3178	3141	3148	3192	3229	3317	3362	3430
Gyneekologie a porodnictví	2527	2513	2509	2512	2471	2417	2423	2414	2402	2388	2389	2396	2401
Klinický psycholog	907	929	941	946	932	903	913	934	949	981	990	1000	1008
Neurologie	1214	1209	1216	1222	1196	1143	1140	1149	1151	1146	1156	1153	1156
Oftalmologie	1124	1128	1124	1128	1108	1083	1086	1104	1101	1100	1103	1103	1109
Ortopedie a traumatologie pohybového ústrojí	1619	1617	1623	1630	1604	1574	1582	1581	1582	1572	1585	1597	1600
Otorinolaryngologie a chirurgie hlavy a krku	927	926	924	929	902	879	879	878	872	865	866	860	859
Praktické lékařství pro děti a dorost	2833	2802	2790	2755	2729	2684	2665	2625	2596	2572	2560	2537	2532
Praktické lékařství	2598	2587	2565	2542	2524	2481	2469	2457	2463	2458	2452	2448	2447
Psychiatrie	1266	1271	1296	1302	1286	1243	1248	1249	1264	1281	1284	1292	1294
Rehabilitační a fyzikální medicína	1489	1495	1489	1488	1478	1434	1426	1406	1409	1410	1406	1411	1428
Vnitřní lékařství	2584	2579	2575	2579	2515	2444	2441	2443	2443	2435	2420	2418	2411
Všeobecná sestra	1454	1444	1547	1601	1601	1591	1633	1644	1661	1706	1750	1784	1824
Všeobecné praktické lékařství	7059	7008	7001	6983	6960	6871	6875	6869	6854	6833	6836	6830	6809
Zubní lékařství	7143	7078	7057	7033	7024	6926	6889	6807	6784	6698	6683	6653	6656
Zubní technik	1895	1884	1893	1883	1882	1851	1847	1833	1828	1820	1818	1813	1823

V levém sloupci je jsou názvy jednotlivých oborů péče. Další sloupce obsahují počty poskytovatelů zdravotnických služeb v oboru péče k danému měsíci. Řádky jsou seřazeny abecedně podle názvu oboru péče.

Struktura upraveného souboru

V rámci čištění dat byla nejdříve provedena detekce krajních hodnot (dále jen outlierů) pro všechny obory péče. Následně se outliery nahradily hodnotami očekávanými na jejich pozicích. Dále byla provedena normalizace a diskretizace vybraných oborů. Těmito obory byly neurologie pro normalizaci a psychiatrie pro diskretizaci, neboť tyto obory neobsahovaly žádné outliery, a tak se pro ukázkou řešení zdály být dobrými kandidáty (jinak je výběr oboru samozřejmě irelevantní).

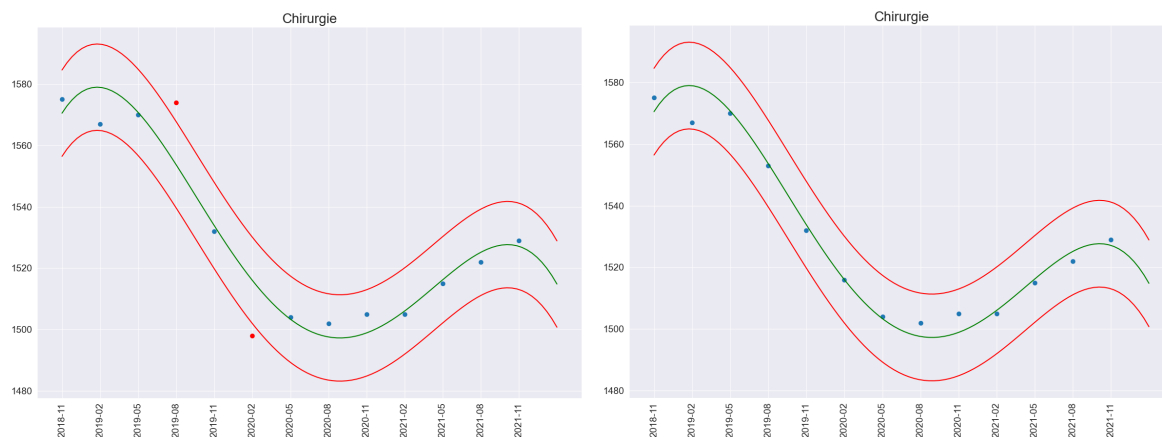
Detekce outlierů

K detekci outlierů byla použita metoda modifikované z-skóre. Jedná se o robustní statistickou metodu založenou na mediánu absolutní odchylky (dále jen MAD), jichž se využívá k detekci anomálií v časových řadách. Princip spočívá v tom, že se každé hodnotě vypočítá její z-skóre, jehož modifikace je následující:

$$Z = \frac{X - \mu}{\sigma} \quad \rightarrow \quad Z = 0.6745 \times \frac{X - \text{median}(X)}{\text{MAD}(X)}$$

Za outlier je považována každá hodnota, která neleží v intervalu $\langle S - z_limit; S + z_limit \rangle$, kde S je nějakým způsobem zvolený střed dat (může být střední hodnota, očekávaná hodnota apod.) a z_limit je předem zvolená anomální hranice.

V naší úloze se za S vždy zvolila očekávaná hodnota získaná polynomiální regresí 4. řádu. Za povolený šum dat se zvolilo normální rozložení kolem regresní křivky. Detekované outliery se následně nahradily hodnotou regresní křivky jak ukazují následující grafy.



Grafy pro všechny obory péče obsahující outliery jsou přiloženy v [příloze A](#). V [příloze B](#) pak jsou výsledky oborů bez outlierů. Celkem bylo detekováno 12 outlierů a to v těchto oborech:

Obor péče

Anesteziologie a intenzivní medicína:

Chirurgie:

Dermatovenerologie:

Endokrinologie a diabetologie:

Gynekologie a porodnictví:

Ortopedie a traumatologie pohybového ústrojí:

Otorinolaryngologie a chirurgie hlavy a krku:

Všeobecná sestra:

Všeobecné praktické lékařství:

Měsíc

2019-08

2019-08

2020-02

2020-02

2020-02

2019-08

2020-02

2020-02

2019-08

2020-02

2019-02

2020-02

Oprava outlieru

975 → 955

1574 → 1553

1498 → 1516

1027 → 1042

943 → 960

2512 → 2491

2417 → 2442

1574 → 1590

929 → 915

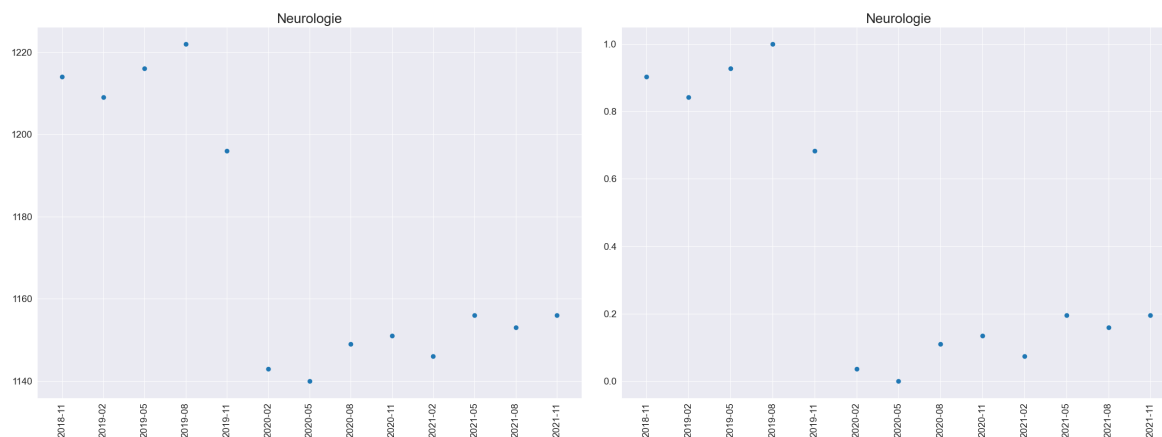
879 → 890

1444 → 1494

6871 → 6904

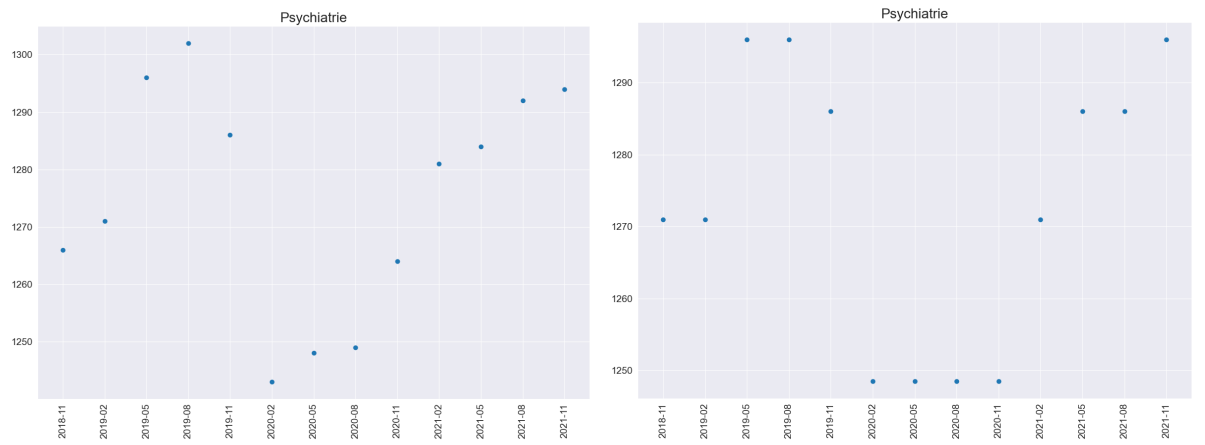
Normalizace

K normalizaci byl zvolen obor neurologie, jehož hodnoty byly převedeny metodou Min-Max do rozsahu $<0, 1>$.



Diskretizace

Diskretizace byla provedena nad oborem psychiatrie rozdělením podle stejné hloubky. Poněvadž má datová sada 13 atributů, rozumný počet košů je 4, tedy každý by měl obsahovat 3 hodnoty a jeden 4. Rozdělení do košů proběhlo podle kvartilů a vyhlazení pak mediánem jednotlivých košů.



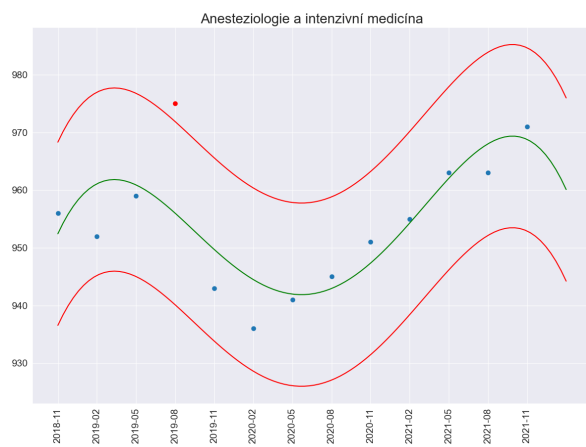
Výsledný csv soubor

Obor péče	2018-11	2019-02	2019-05	2019-08	2019-11	2020-02	2020-05	2020-08	2020-11	2021-02	2021-05	2021-08	2021-11
Anesteziologie a intenzivní medicína	956.0	952.0	959.0	955.0	943.0	936.0	941.0	945.0	951.0	955.0	963.0	963.0	971.0
Chirurgie	1575.0	1567.0	1570.0	1553.0	1532.0	1516.0	1504.0	1502.0	1505.0	1505.0	1515.0	1522.0	1529.0
Dermatovenerologie	1069.0	1073.0	1076.0	1068.0	1056.0	1042.0	1031.0	1031.0	1030.0	1016.0	1016.0	1019.0	1026.0
Endokrinologie a diabetologie	994.0	994.0	998.0	1000.0	977.0	960.0	946.0	945.0	949.0	937.0	930.0	930.0	931.0
Fyzioterapeut	3045.0	3072.0	3115.0	3146.0	3177.0	3178.0	3141.0	3148.0	3192.0	3229.0	3317.0	3362.0	3430.0
Gynekologie a porodnictví	2527.0	2513.0	2509.0	2491.0	2471.0	2442.0	2423.0	2414.0	2402.0	2388.0	2389.0	2396.0	2401.0
Klinický psycholog	907.0	929.0	941.0	946.0	932.0	903.0	913.0	934.0	949.0	981.0	990.0	1000.0	1008.0
Neurologie	0.90243902	0.84146341	0.92682926	1.0	0.68292682	0.03658536	0.0	0.10975609	0.13414634	0.07317073	0.19512195	0.15853658	0.19512195
Oftalmologie	1124.0	1128.0	1124.0	1128.0	1108.0	1083.0	1086.0	1104.0	1101.0	1100.0	1103.0	1103.0	1109.0
Ortopedie a traumatologie pohybového ústrojí	1619.0	1617.0	1623.0	1630.0	1604.0	1590.0	1582.0	1581.0	1582.0	1572.0	1585.0	1597.0	1600.0
Otorinolaryngologie a chirurgie hlavy a krku	927.0	926.0	924.0	915.0	902.0	890.0	879.0	878.0	872.0	865.0	866.0	860.0	859.0
Praktické lékařství pro děti a dorost	2833.0	2802.0	2790.0	2755.0	2729.0	2684.0	2665.0	2625.0	2596.0	2572.0	2560.0	2537.0	2532.0
Praktické lékařství	2598.0	2587.0	2565.0	2542.0	2524.0	2481.0	2469.0	2457.0	2463.0	2458.0	2452.0	2448.0	2447.0
Psychiatrie	1271.0	1271.0	1296.0	1296.0	1286.0	1248.5	1248.5	1248.5	1248.5	1271.0	1286.0	1286.0	1296.0
Rehabilitační a fyzikální medicína	1489.0	1495.0	1489.0	1488.0	1478.0	1434.0	1426.0	1406.0	1409.0	1410.0	1406.0	1411.0	1428.0
Vnitřní lékařství	2584.0	2579.0	2575.0	2579.0	2515.0	2444.0	2441.0	2443.0	2443.0	2435.0	2420.0	2418.0	2411.0
Všeobecná sestra	1454.0	1494.0	1547.0	1601.0	1601.0	1591.0	1633.0	1644.0	1661.0	1706.0	1750.0	1784.0	1824.0
Všeobecné praktické lékařství	7059.0	7008.0	7001.0	6983.0	6960.0	6904.0	6875.0	6869.0	6854.0	6833.0	6836.0	6830.0	6809.0
Zubní lékařství	7143.0	7078.0	7057.0	7033.0	7024.0	6926.0	6889.0	6807.0	6784.0	6698.0	6683.0	6653.0	6656.0
Zubní technik	1895.0	1884.0	1893.0	1883.0	1882.0	1851.0	1847.0	1833.0	1828.0	1820.0	1818.0	1813.0	1823.0

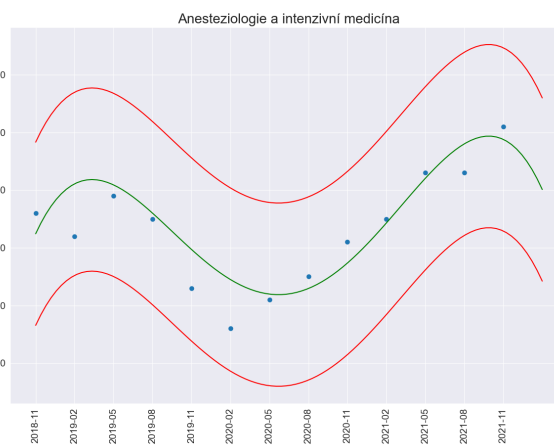
Pozn. Jak lze vidět, úprava dat probíhala pro jednotlivé časové řady oborů, nikoliv pro atributy (sloupce). Pro toto konkrétní zadání vzhledem k povaze časových řad to dává větší smysl, než napříč řadami pro atributy (postup schválen na fóru).

Příloha A

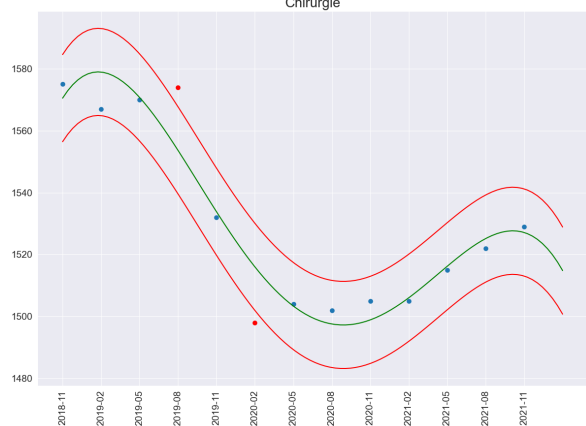
S outliery



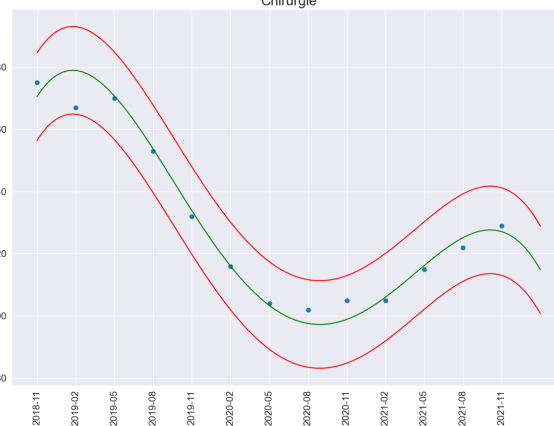
Po vyčištění



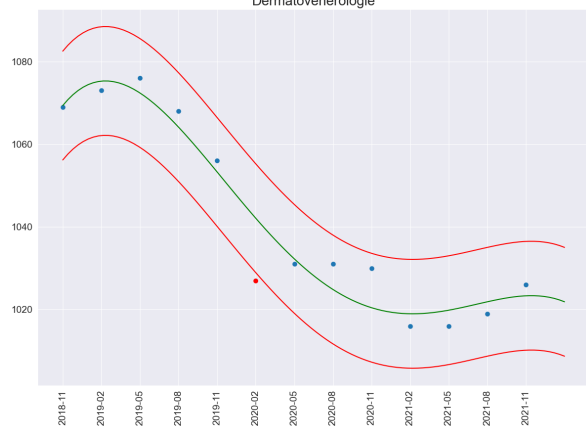
Chirurgie



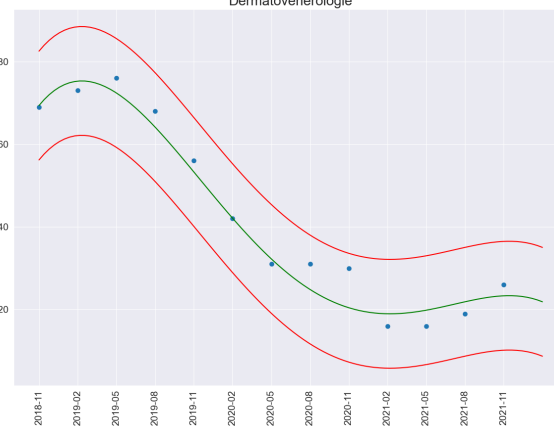
Chirurgie

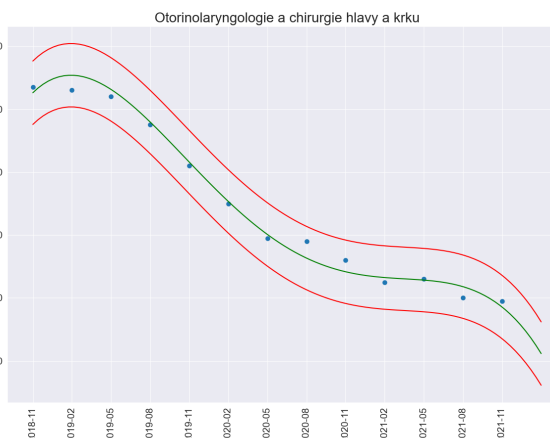
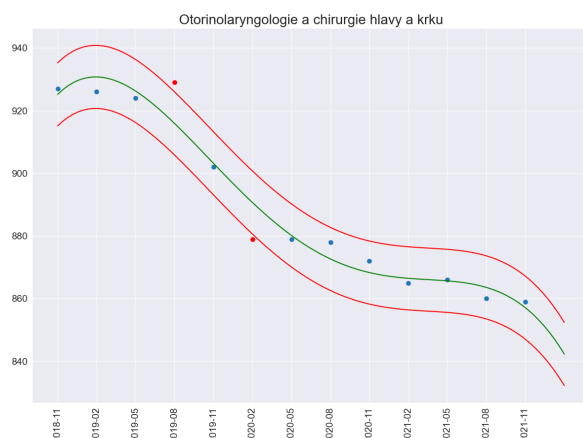
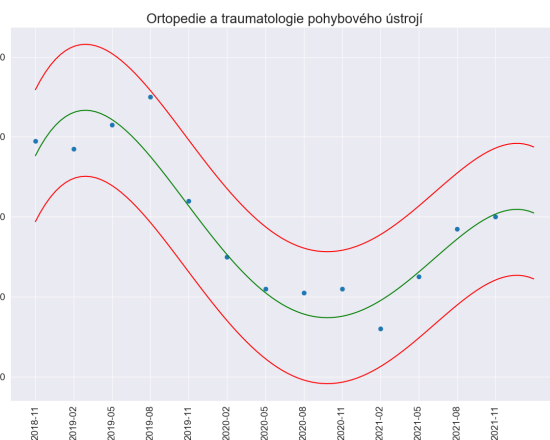
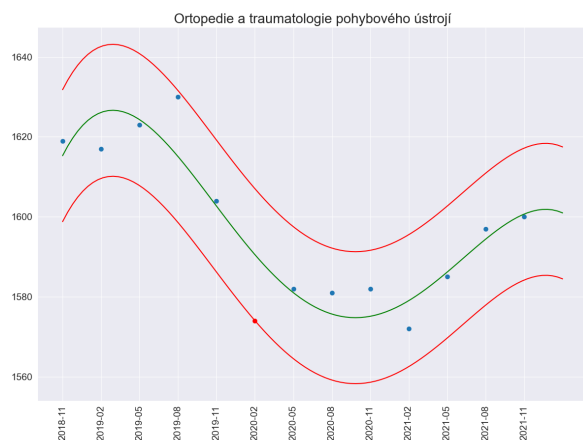
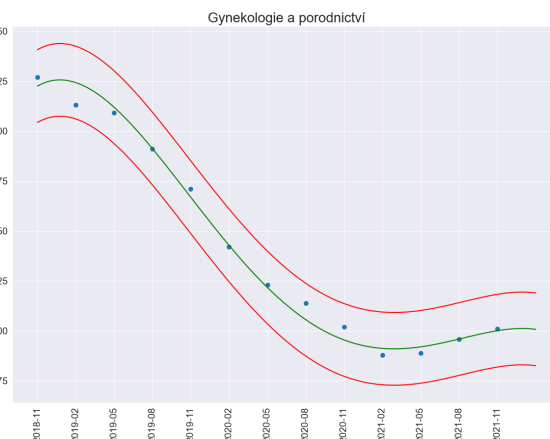
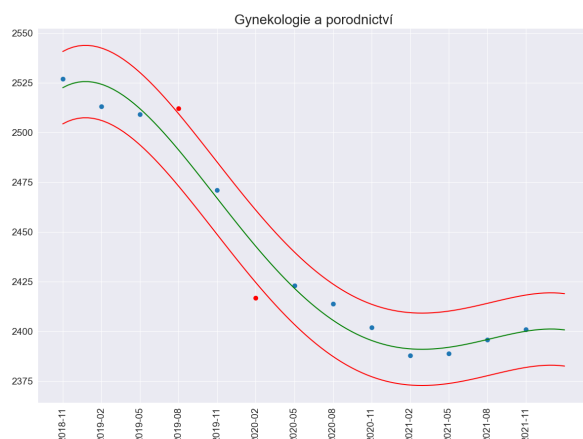
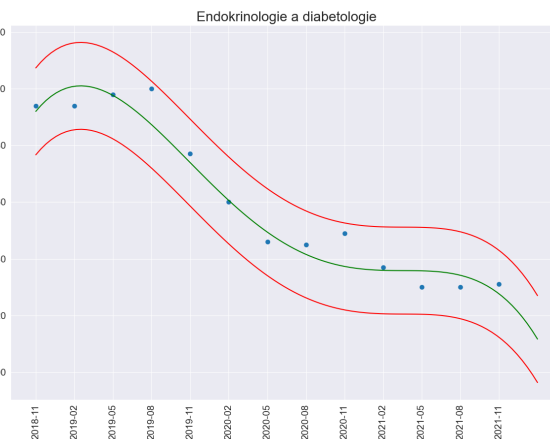
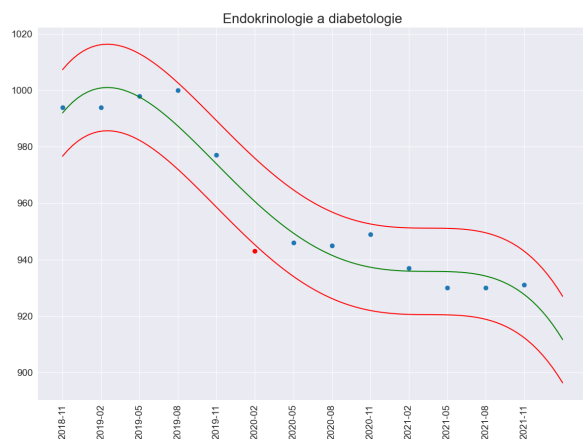


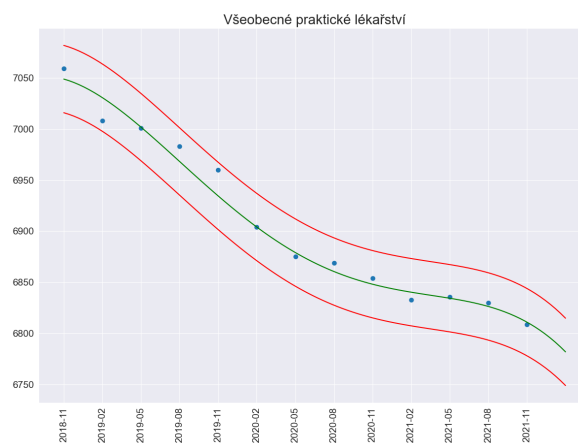
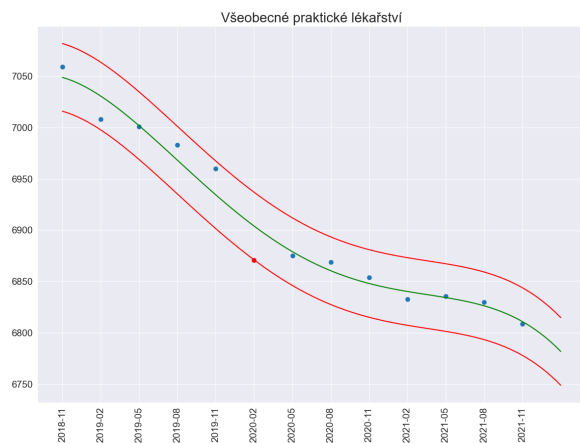
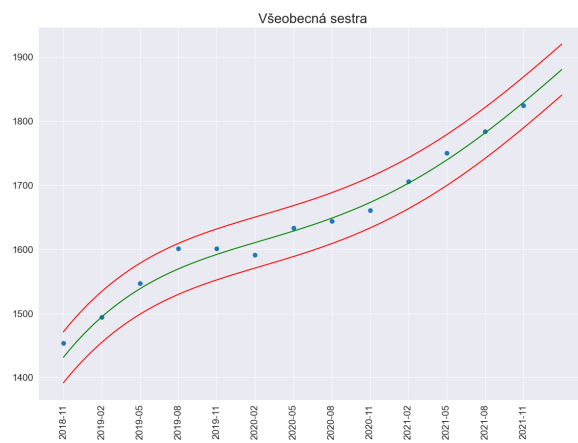
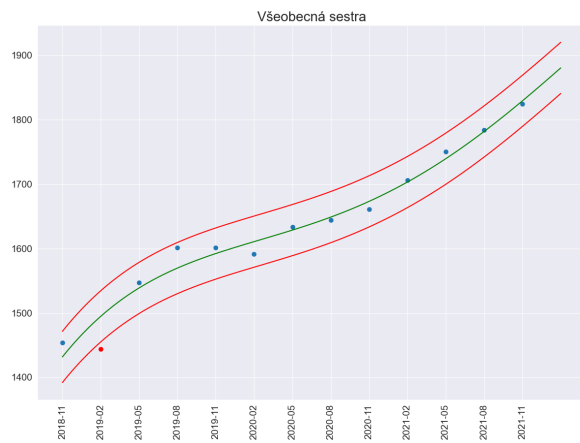
Dermatovenerologie



Dermatovenerologie







Příloha B

