

[Click here to toggle on/off the raw code.](#)

Návrh vizualizací a reportů pro MHMP

Jakub Vöröš

První soubor 'kontejnery.csv'

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 440 entries, 0 to 439
Data columns (total 16 columns):
id                440 non-null int64
code              440 non-null object
latitude          440 non-null object
longitude         440 non-null object
address           440 non-null object
district          27 non-null object
postal_code       440 non-null int64
total_volume      440 non-null int64
trash_type        440 non-null object
prediction        283 non-null object
bin_type          424 non-null object
installed_at      418 non-null object
created           440 non-null object
last_modify       440 non-null object
created_batch_id  440 non-null int64
updated_batch_id  440 non-null int64
dtypes: int64(5), object(11)
memory usage: 55.1+ KB
```

True

True

440

Přehled důležitých úprav a sloupků

'id' - unikátní číselné označení

'code' - složen z kódů pro umístění začínajícího n 'S' a kontejneru na 'C'; upraveny za pomoci dalších tabulek

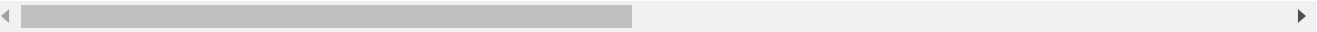
'latitude', 'longitude' - nahrazeny ',' za '.', přeměna na float

'district' - pouze 27; hodnoty 'Praha 1', 'TEST', 'Kladno'; řádky s údaji obsahující 'Kladno' odstraněny, 'TEST' nahrazen null, doplněny hodnoty díky porovnání Point a Polygon

NaN 413
Kladno 16
TEST 6
Praha 1 5
Name: district, dtype: int64

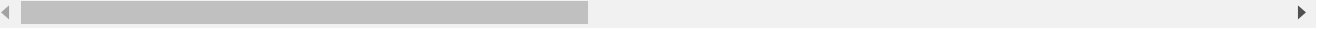
13000 82
11000 78
19000 75
18000 57
16000 48
15300 36
10000 27
27201 16
15000 13
17000 8
Name: postal_code, dtype: int64

	id	code	latitude	longitude	address	district	postal_code	total_volume	trash_ty
63	29868	0001/ 260C00161	50.089248	14.346298	Čílova 281	TEST	16000	4000	pař
64	29869	0001/ 260C00162	50.089248	14.346298	Čílova 281	TEST	16000	4000	plas
65	29870	0001/ 260C00163	50.089248	14.346298	Čílova 281	TEST	16000	1500	pař
66	29871	0001/ 260C00164	50.089248	14.346298	Čílova 281	TEST	16000	1500	beverage_carto
67	29872	0001/ 260C00165	50.089248	14.346298	Čílova 281	TEST	16000	1500	glass_colour
68	29873	0001/ 260C00166	50.089248	14.346298	Čílova 281	TEST	16000	1500	glass_wh



NaN 419
Praha 1 5
Name: district, dtype: int64

	OBJECTID	DAT_VZNIK	DAT_ZMENA	PLOCHA	ID	KOD_MC	NAZEV_MC	KOD_MO	KOD_
0	1	20181106141412	20181106164427	10183715.85	25	547310	Praha- Čakovice	94	



```
Praha 1      83
Praha 3      82
Praha 9      70
Praha 8      53
Praha 6      48
Praha 16     36
Praha 10     27
Praha 18      9
Praha 7       8
Praha 5       8
Name: district, dtype: int64
```

```
paper      102
plastic     98
beverage_cartons  68
glass_coloured  64
glass_white   62
metal       30
Name: trash_type, dtype: int64
```

```
Bammens Memphis      161
Schäfer/Europa-0R/GB  122
Schäfer/Europa-0V     74
Nadzem                43
Bammens Belfast       18
Meva Top System        4
MINI H-A               2
Name: bin_type, dtype: int64
```

```
3000    214
1500    190
1100     8
2000     6
4000     6
Name: total_volume, dtype: int64
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 424 entries, 16 to 439
Data columns (total 17 columns):
id                424 non-null int64
code              424 non-null object
latitude          424 non-null float64
longitude         424 non-null float64
address           424 non-null object
district          424 non-null object
postal_code       424 non-null int64
total_volume      424 non-null int64
trash_type        424 non-null object
prediction         276 non-null datetime64[ns]
bin_type          424 non-null object
installed_at      418 non-null datetime64[ns]
created           424 non-null datetime64[ns]
last_modify       424 non-null datetime64[ns]
created_batch_id  424 non-null int64
updated_batch_id  424 non-null int64
geometry          424 non-null object
dtypes: datetime64[ns](4), float64(2), int64(5), object(6)
memory usage: 59.6+ KB
```

Druhý soubor 'mereni.csv'

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 268297 entries, 0 to 268296
Data columns (total 16 columns):
id                268297 non-null int64
container_id      268297 non-null int64
code              268297 non-null object
percent_calculated 268297 non-null object
upturned          268297 non-null int64
temperature       268297 non-null object
battery_status    268297 non-null object
measured_at       268297 non-null object
measured_at_utc   268297 non-null object
prediction        175666 non-null object
prediction_utc     175666 non-null object
firealarm         268297 non-null int64
created           268297 non-null object
last_modify       261703 non-null object
created_batch_id  268297 non-null int64
updated_batch_id  261703 non-null float64
dtypes: float64(1), int64(5), object(10)
memory usage: 32.8+ MB
```

True

False

Přehled důležitých úprav a sloupků

- 'id' - unikátní číselné označení pro každé měření
- 'container_id' - stejná jako z předchozí tabulky; 29928 - 895x, 29895 - 876x, 30143 - 118x, 30144 - 85x; unikátních 424
- 'code' - viz předešlá tabulka; S0024C00243 - 694x, S0024C00242 - 677x; unikátních 850; loop vytvořen slovník s chybnými a správnými kódy pro opravu; 5 špatných
- 'percent_calculated' - nahrazeny ',' za '.', přeměna na float
- 'upturned' - 3639x povalena

```
30143    118
30144     85
Name: container_id, dtype: int64
```

```
S0024C00243    694
S0024C00242    677
Name: code, dtype: int64
```

850

424

Jelikož máme 424 unikátních kontejnerových id v tabulkách 'kont' a 'mer', víme, že toto číslo sedí.

Nyní se budeme snažit zjistit proč máme 850 kódových označení a vybrat ty správná.

Nyní máme 419 koretních kódových označení a 5 špatných, které nelze dohledat ani v třetí tabulce.

```
0      264658
1       3639
Name: upturned, dtype: int64
```

```
0      268297
Name: firealarm, dtype: int64
```

Z dat jsme se dozvěděli, že z více jak 268 tisíc měření byla přibližně 3600 krát popelnice povalena, ale ani jednou nebyl zaznamenán požár.

Třetí soubor 'svozy.csv'

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8697 entries, 0 to 8696
Data columns (total 12 columns):
id                8697 non-null int64
container_id      8697 non-null int64
code              8697 non-null object
pick_at           8697 non-null object
pick_at_utc       8697 non-null object
percent_before    8697 non-null object
percent_now       8697 non-null object
event_driven      8697 non-null int64
created           8697 non-null object
last_modify       6534 non-null object
created_batch_id  8697 non-null int64
updated_batch_id  6534 non-null float64
dtypes: float64(1), int64(4), object(7)
memory usage: 815.5+ KB
```

```
30062      1
30352      1
Name: container_id, dtype: int64
```

```
S0023C00234      48
S0013C00131      48
Name: code, dtype: int64
```

Přehled důležitých úprav a sloupků

- 'id' - unikátní číselné označení pro každý svoz
- 'container_id' - 29904 - 63x, 29908 - 62x, 30062 - 1x, 30352 - 1x
- 'code' - viz předešlá tabulka; S0023C00234 - 48, S0013C00131 - 48x
- 'percent_before' - nahrazeny ';' za '.', přeměna na float
- 'percent_now' - nahrazeny ';' za '.', přeměna na float; 47x 100%
- 'event_driven' - 0x

Již teď si můžeme všimnout, že zde pracujeme s daleko menším množstvím měření, ale ani zde tyto dva sloupky nesedí

```
0      8697
Name: event_driven, dtype: int64

count      8697.000000
mean        67.581580
std         30.436854
min          0.000000
25%         44.000000
50%         72.000000
75%         99.000000
max        100.000000
Name: percent_before, dtype: float64

count      8697.000000
mean        14.329884
std         20.417364
min          0.000000
25%          0.000000
50%          6.000000
75%         20.000000
max        100.000000
Name: percent_now, dtype: float64
```

47

	code	latitude	longitude	address	district	postal_code	total_volume	trash_ty
id								
29786	S0020C00201	50.102927	14.368419	Na Pernikářce 734	Praha 6	16000	1500	beverage_carto
29787	S0020C00202	50.102927	14.368419	Na Pernikářce 734	Praha 6	16000	1500	glass_colour

Analýza

Úkolem je vytvořit návrh vizualizace a reportů pro MHMP, které napomůžou s rozhodováním v rámci odpadového hospodářství.

Prohlídka top 10 kontejnerů s největším množstvím měření

total_frequency	
container_id	
29928	895
29895	876
29927	873
30074	793
29892	792
30090	791
29894	790
30073	790
30034	789
30071	789

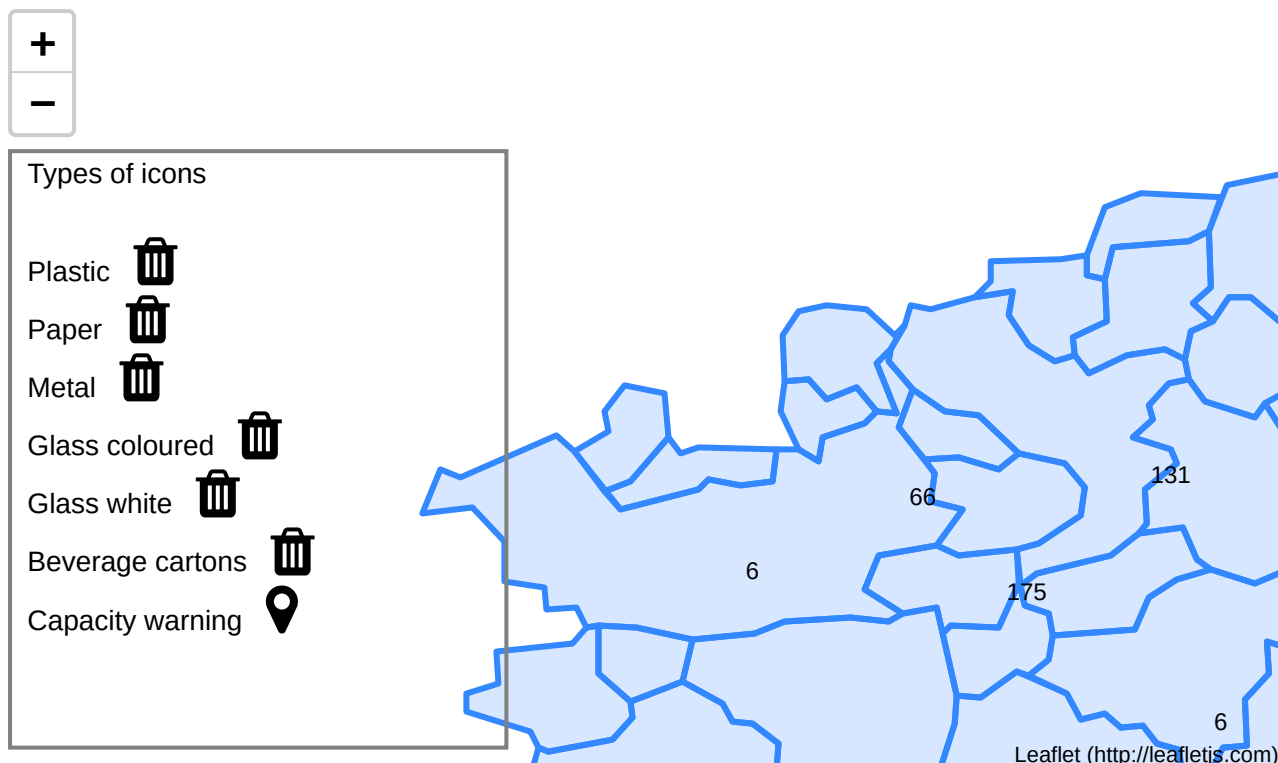
Celkový počet unikátních měření je 268297

Vytvoření nové tabulky, která nás seznámí s různými druhy četnosti zaplnění jednotlivých kontejneru získané z dat tabulky 'mereni'.

	total_frequency	freq_cont_fullness_>50%	freq_cont_fullness_>75%	freq_cont_fullness_>99%
container_id				
29793	750	740.0	739.0	687.0
29834	753	743.0	723.0	621.0
29974	732	718.0	687.0	64.0
29932	751	695.0	663.0	342.0
29954	703	667.0	611.0	289.0
30162	499	466.0	425.0	63.0
29882	774	715.0	642.0	273.0
29835	739	676.0	609.0	368.0
29897	788	697.0	634.0	340.0
30120	652	575.0	524.0	108.0



Nyní promítnemé zjištěné informace do vizuální podoby. Jako vhodné se jeví interaktivní znázornění ve formě mapy s informačním textem.



Mapa se zdá být přehledným a snadným znázorněním informačních údajů získaných z měření. Všechny kontejnery lze opticky přiřadit pod konkrétní čtvrť, jejíž název se zobrazí po najetí myši. Jednotlivé ikony jsou při originálním oddálení zobrazeny ve shluku a tak se vizuálně nezastiňují a tak mapa nepůsobí hekticky. Po přiblížení se ikony rozprostřou na konkrétní lokace a po rozkliknutí se zobrazí přesný počet a barvou zabarvený typ kontejnerů. Po druhém kliknutí již na konkrétní jednu ikonu se zobrazí tabulka s popisem. V tomto provedení je vše tabulky id kontejneru, procentem zobrazená četnost naplnění popelnice z více jak 75%, průměrná zaplněnost odpadem též v procentech, objem nádoby a pokud se jedná o konkrétní kontejner, který je často přeplněn, je zde i na začátku tabulky na toto upozornění a samotná ikona na mapě je i rudě ohraničena pro rychlou identifikaci.

Jelikož se dokumentace a i tabulky zaměřují na zaplněnost ze 75%, budeme s touto hodnout počítat i v dalších analýzách

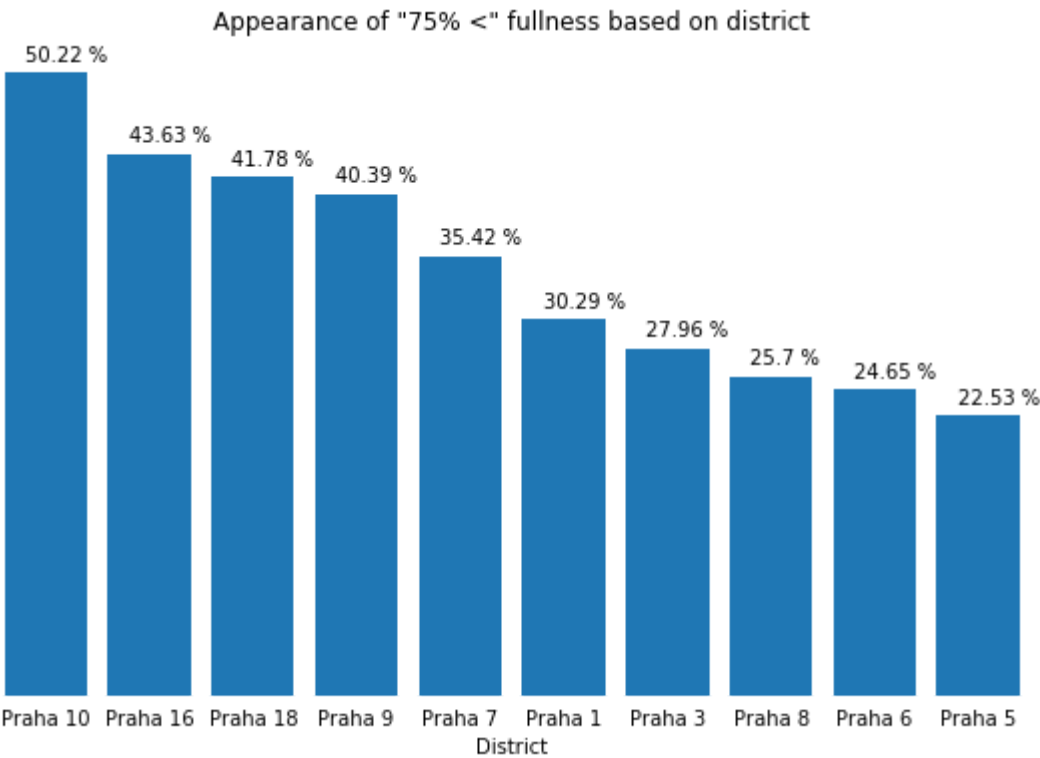
DATA >75%

32.77% ze všech měření obsahuje údaje zaplněnosti vyšších než 75% kapacity a 44.1% naopak nižších než 50%.

	trash_type	beverage_cartons	glass_coloured	glass_white	metal	paper	plastic	All
district	total_volume							
Praha 1	1100	67.89	23.85	0.00	8.26	0.00	0.00	218
	1500	43.74	17.32	0.14	38.81	0.00	0.00	1461
	2000	0.00	0.00	0.00	0.00	64.70	35.30	762
	3000	5.61	0.20	0.18	4.01	55.63	34.36	14715
Praha 10	1100	0.00	0.00	0.00	100.00	0.00	0.00	434
	1500	36.97	16.27	13.15	0.00	8.30	25.31	3338
	3000	6.02	0.00	0.00	0.00	36.41	57.57	4897
Praha 16	3000	20.39	0.00	0.00	12.08	33.27	34.25	8872
Praha 18	1500	0.00	80.54	19.46	0.00	0.00	0.00	406
	3000	19.56	0.00	0.00	0.00	34.98	45.46	1641
Praha 3	1500	20.27	10.16	6.22	0.00	12.62	50.73	10915
	3000	2.46	1.42	0.00	0.00	83.63	12.49	4874
Praha 5	1500	0.00	100.00	0.00	0.00	0.00	0.00	21
	3000	0.00	0.00	0.00	0.00	46.38	53.62	1175
Praha 6	1500	23.79	43.25	31.28	0.00	1.68	0.00	2858
	3000	13.62	0.00	1.10	0.00	50.65	34.63	3826
	4000	0.00	0.00	0.00	0.00	61.19	38.81	1309
Praha 7	1500	0.00	79.03	20.97	0.00	0.00	0.00	434
	3000	0.00	0.00	0.00	0.00	53.92	46.08	1606
Praha 8	1500	22.06	29.71	28.81	16.24	3.19	0.00	2888
	3000	0.00	0.00	0.00	0.00	59.05	40.95	3734
Praha 9	1100	0.00	0.00	0.00	100.00	0.00	0.00	145
	1500	8.44	39.74	51.82	0.00	0.00	0.00	3815
	3000	24.87	0.00	0.00	0.00	35.89	39.24	13579
All		14.95	7.23	5.76	3.75	35.36	32.96	87923

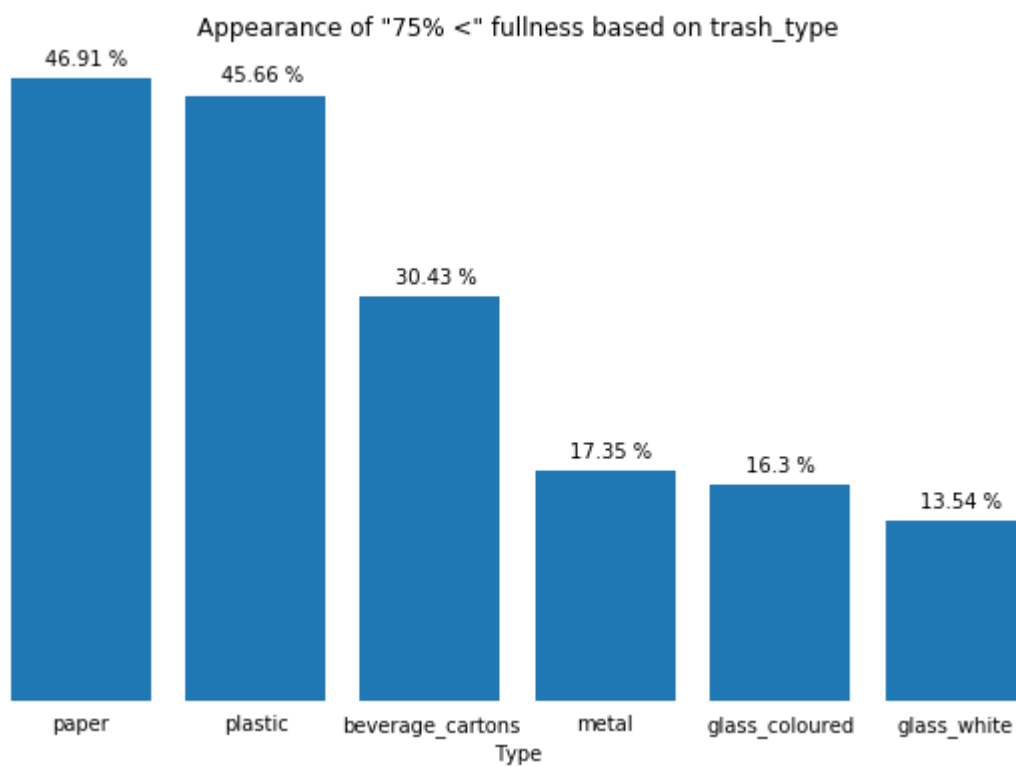
Čtvrti >75%

	75%<_count	75%<_count_in%
district		
Praha 10	8669	50.22
Praha 16	8872	43.63
Praha 18	2047	41.78
Praha 9	17539	40.39
Praha 7	2040	35.42
Praha 1	17156	30.29
Praha 3	15789	27.96
Praha 8	6622	25.70
Praha 6	7993	24.65
Praha 5	1196	22.53



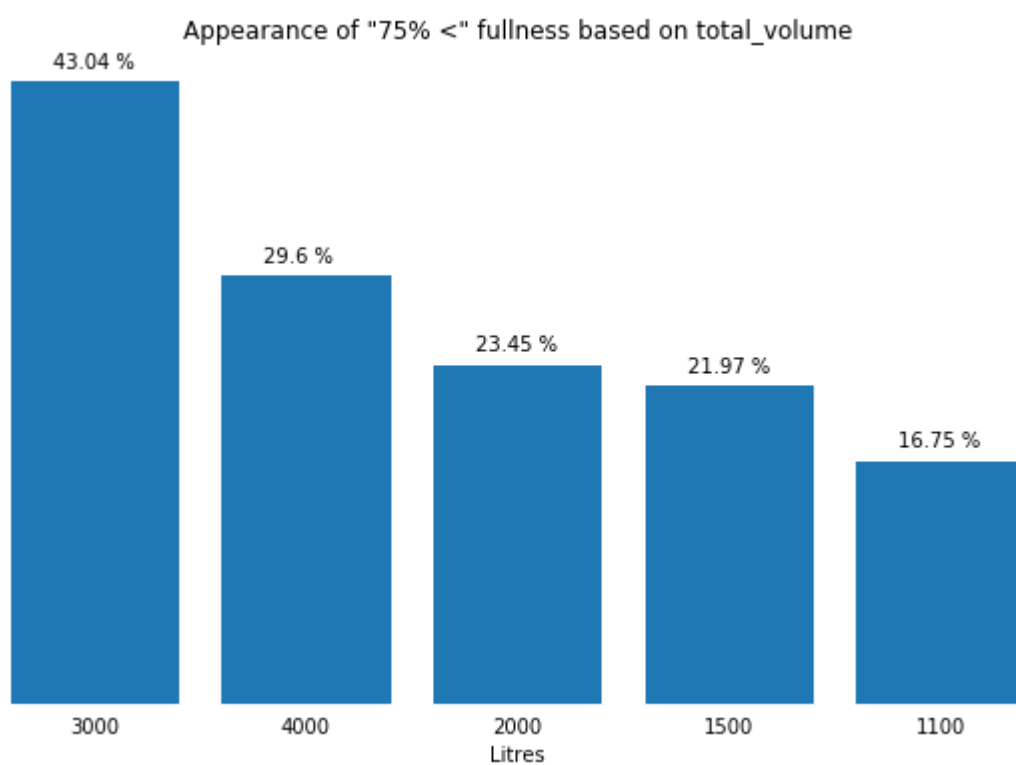
Odpad >75%

	75%<_count	75%<_count_in%
trash_type		
paper	31086	46.91
plastic	28981	45.66
beverage_cartons	13142	30.43
metal	3295	17.35
glass_coloured	6357	16.30
glass_white	5062	13.54



Objem >75%

	75%<_count	75%<_count_in%
total_volume		
3000	58919	43.04
4000	1309	29.60
2000	762	23.45
1500	26136	21.97
1100	797	16.75



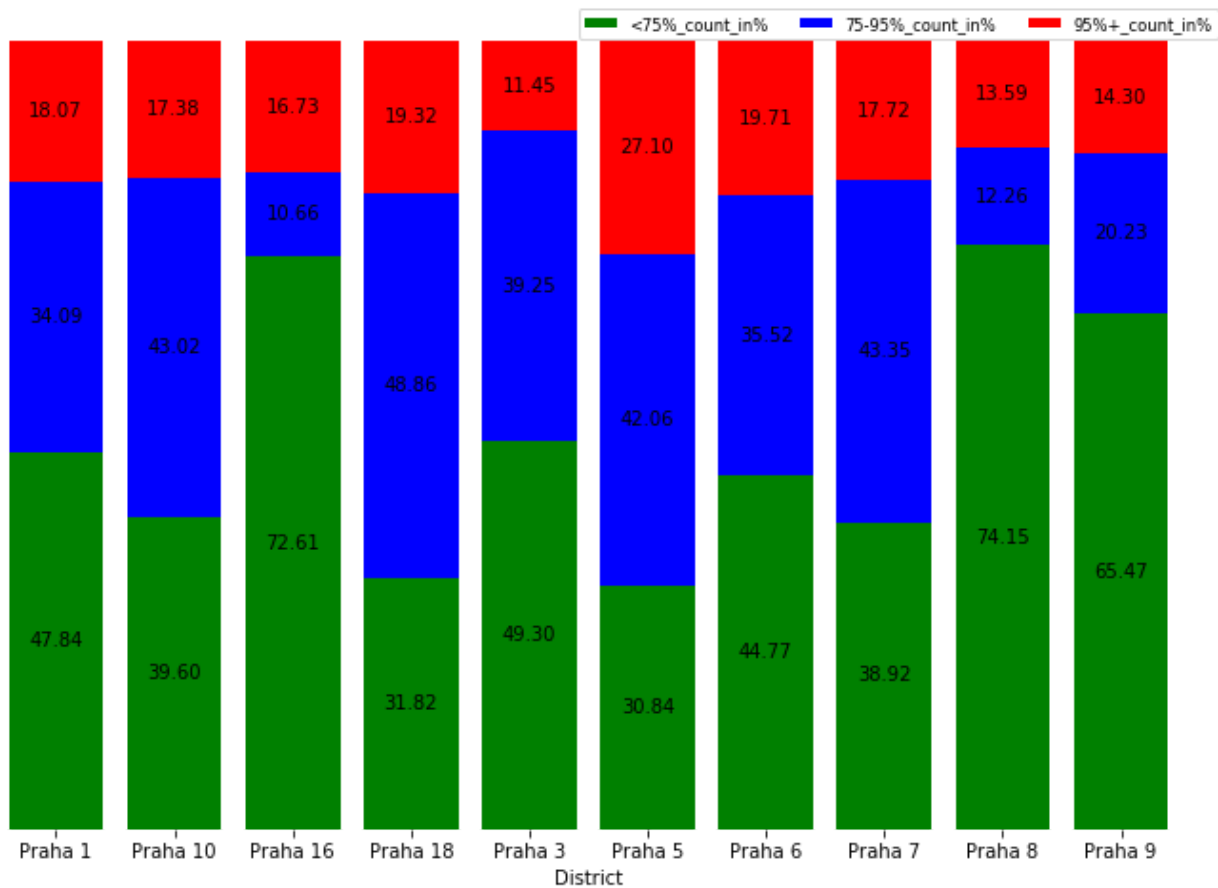
Skutečný stav zaznamenaný při svozu odpadu z tabulky 'svoz'

	Not necessary (<75%)	Overdue (95%+)	Ready to empty (75-95%)	average_percent_before
container_id				
29786	11	0	0	53.91
29787	11	0	0	58.91
29788	5	12	5	89.27
29789	39	11	11	69.25
29790	4	1	0	71.80
29791	0	4	0	100.00
29792	3	8	5	86.00
29793	0	2	0	100.00
29794	0	2	0	98.00
29795	5	5	0	66.10

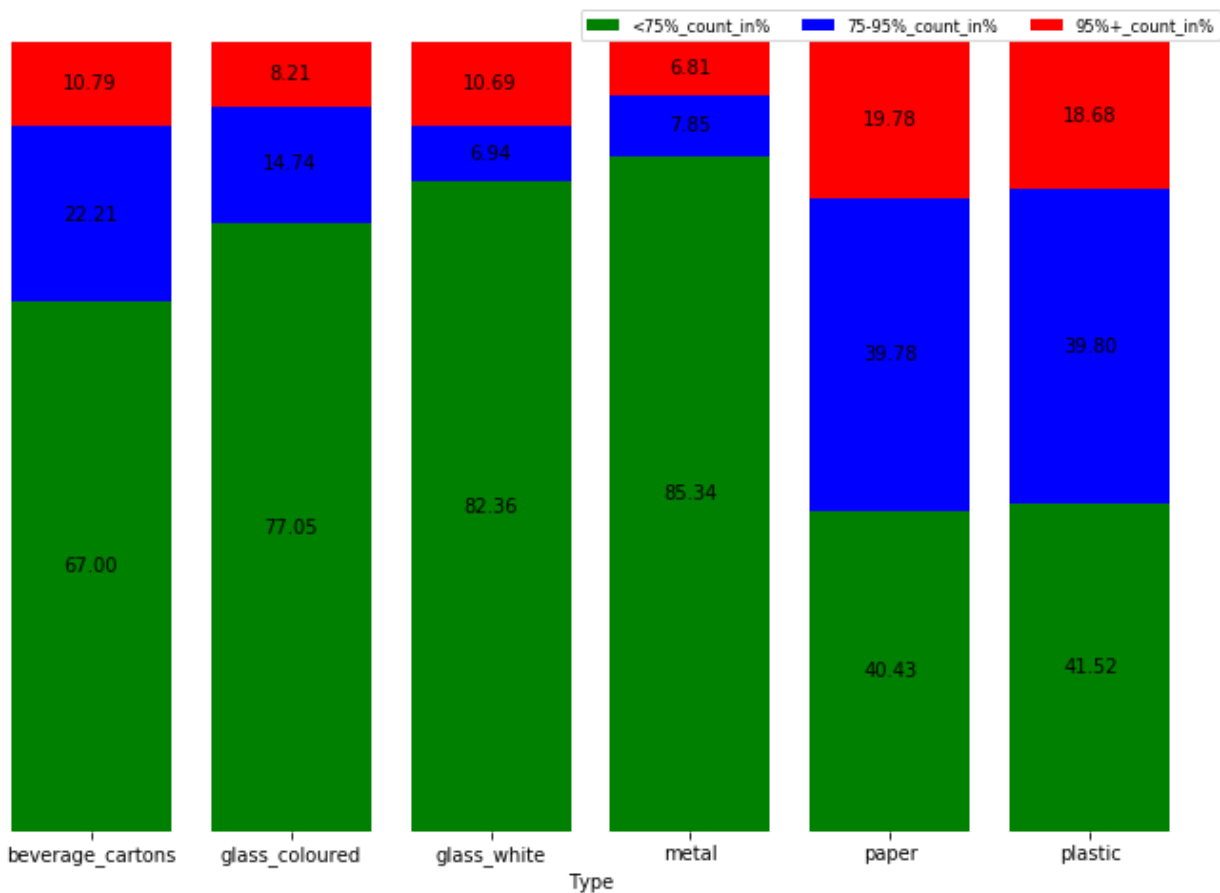
	Not necessary (<75%)	Overdue (95%+)	Ready to empty (75- 95%)	Total	<75%_count_in%	75- 95%_count_in%	95%+_count_in%
district							
Praha 1	1051.0	749.0	397.0	2197.0	47.84	34.09	18.07
Praha 10	139.0	151.0	61.0	351.0	39.60	43.02	17.38
Praha 16	395.0	58.0	91.0	544.0	72.61	10.66	16.73
Praha 18	28.0	43.0	17.0	88.0	31.82	48.86	19.32
Praha 3	956.0	761.0	222.0	1939.0	49.30	39.25	11.45
Praha 5	66.0	90.0	58.0	214.0	30.84	42.06	27.10
Praha 6	552.0	438.0	243.0	1233.0	44.77	35.52	19.71
Praha 7	123.0	137.0	56.0	316.0	38.92	43.35	17.72
Praha 8	611.0	101.0	112.0	824.0	74.15	12.26	13.59
Praha 9	618.0	191.0	135.0	944.0	65.47	20.23	14.30

	Not necessary (<75%)	Overdue (95%+)	Ready to empty (75- 95%)	Total	<75%_count_in%	75- 95%_count_in%	95%+_count_in%
trash_type							
beverage_cartons	540.0	179.0	87.0	806.0	67.00	22.21	10.7
glass_coloured	685.0	131.0	73.0	889.0	77.05	14.74	8.2
glass_white	593.0	50.0	77.0	720.0	82.36	6.94	10.6
metal	326.0	30.0	26.0	382.0	85.34	7.85	6.8
paper	1308.0	1287.0	640.0	3235.0	40.43	39.78	19.7
plastic	1087.0	1042.0	489.0	2618.0	41.52	39.80	18.6

Appearance of fullness based on District



Appearance of fullness based on Type



Na základě všech měření je 52.47% vysypávání odpadu předběžných, kontejnery obsahují < 75%. 31.43% je v rámci mezí, naplněnost je 75-95% a 16.09% je pozdních, nádoby jsou přeplněné

Finální tabulka s porovnáním analýz z "mereni" a "svoz"

container_id	Not necessary (<75%)	Overdue (95%+)	Ready to empty (75- 95%)	average_percent_before	Total	%_occurrence_of_>75%	fullness
29793	0.00	100.00	0.00	100.00	2	98.53	
29834	0.00	100.00	0.00	100.00	2	96.02	
29974	16.67	16.67	66.67	79.67	6	93.85	
29932	25.00	65.00	10.00	77.35	20	88.28	
29954	60.00	40.00	0.00	60.80	10	86.91	
30162	50.00	25.00	25.00	69.50	4	85.17	
29882	33.33	48.15	18.52	74.19	27	82.95	
29835	16.67	83.33	0.00	91.25	12	82.41	
29897	8.70	69.57	21.74	90.09	23	80.46	
30120	50.00	25.00	25.00	75.25	4	80.37	

33

409

Při porovnání tabulek, které vznikly z průběžných měření "mereni" a při svozech "svoz" nám vyplynulo, že 33 kontejnerů má v obou případech vysokou hodnotu nad 75% zaplněnosti objemu a tak by bylo možné je svázat čteněji.