

Big data a small data

v rámci projektu *Nová mobilita – vysokorychlostní dopravní systémy a dopravní chování populace*

Martin Kvizda

II. seminář k vysokorychlostní železnici v České republice
Datová základna a perspektivy jejího využití
Praha, 29. 1. 2020



Cíl

Představit datovou základnu projektu:

- obsah, rozsah a struktura,
- předpokládané využití.

Obsah prezentace

1. Zdroje informací – potřeby a trendy
2. Big data
3. Small data

1. Zdroje informací – potřeby a trendy

- **Four Stage Model (FSTM, 4SM)** - Mitchell a Rapkin (1954):
 1. identifikace zdroje cest (trip generation),
 2. rozdělení cest (trip distribution),
 3. volba dopravního módu (mode choice),
 4. identifikace tras (origin – destination matrix).
- Roth (2009): 1990s v USA růst přepravní kapacity dálnic **7x pomaleji, než poptávka** po přepravě;
- Flyvbjerg (2006): 210 projektů ŽDI, 14 zemí, celkem 58 mld. USD: **90 % predikce poptávky po přepravě nadhodnocena** (72 % nadhodnocení vyšší než 60 %).
- Martens (2016): posledních 50 let dopravní politiky a plánování = infrastruktura sama o sobě **bez zájmu o ty, kdo tento systém používají.**

➤ **Small Data:**

- informace o vzorku respondentů omezeném výběrem a časem,
- spotřebitelské šetření (dotazníkový průzkum, sčítací kampaň, rozhovor...),
- široké spektrum detailních informací,
- vypovídací schopnost omezená velikostí vzorku.

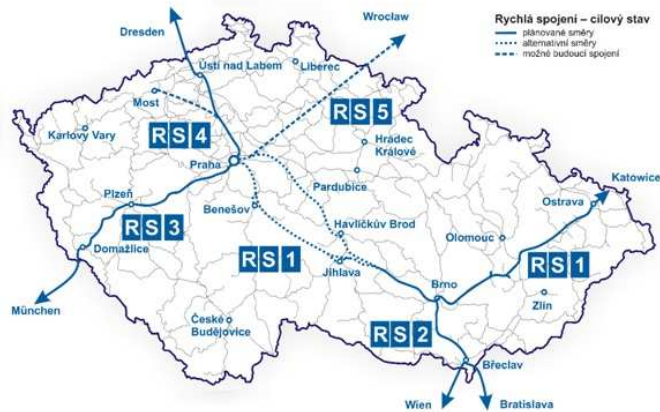
➤ **Big Data:**

- informace o všech cestovatelích za dlouhý časový úsek,
- zbytková signalizační data mobilních operátorů,
- pouze základní kategorie informací,
- vysoká vypovídací schopnost.

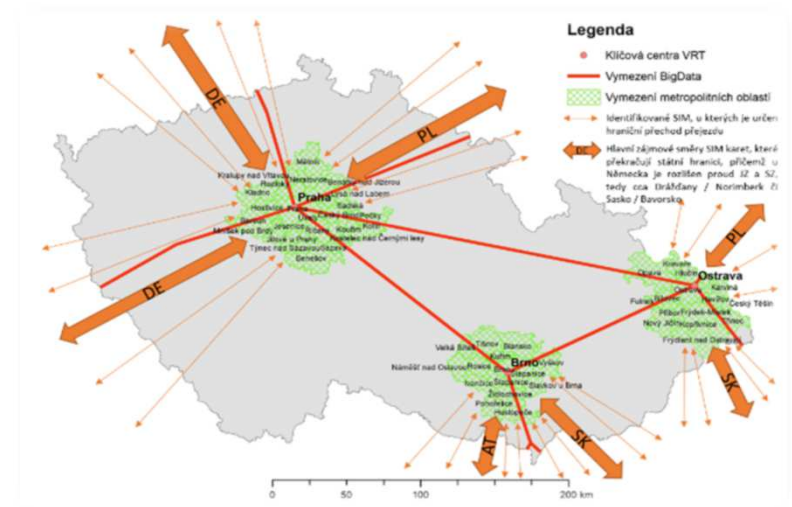
„*Trajectory data rich but activity information poor problem*“ (Gong et al. 2015, Chen et al. 2016):

➤ **kombinace *big a small data***

plán:



realita ???



- skutečná poptávka po přepravě – velikost a struktura,
- motivace cestovatelů – potenciál změny poptávky,
- predikce poptávky po nové dopravní službě – VRT.

2. Big data

- zbytková signalizační data mobilních operátorů
- veřejná zakázka:
 - tržní konzultace,
 - oslovení dodavatelů,
 - best value approach.

- dodavatel:



- dodávka: únor – prosinec 2019

Objem, rozsah a formát dat

- **4 období sběru** = 4 datové sady á 14 dní:
 - Muni4 – duben 2019
 - Muni5 – přelom května a června
 - Muni7 – přelom července a srpna
 - Muni9 – říjen 2019

- **formát dat:** csv / xlsx / parq

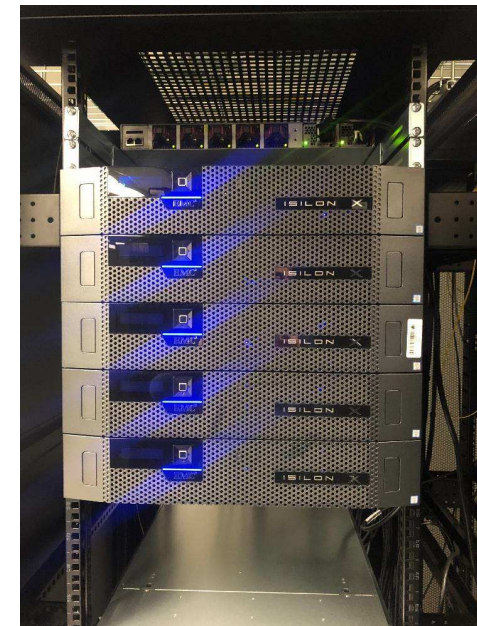
- 1 datová sada = 314 497 souborů (3x formát = 943 491 souborů)
- **4 období = 4 sady = 1 257 988 souborů** (3x formát celkem 3 773 964 souborů)

➤ **úložiště:**

- majetek MUNI,
- clusterové úložiště NAS EMC Isilon (110TB) – záloha na ČD,
- přístupné partnerům projektu,
- celkový objem dat 12 TB.

➤ **způsob zpracování:**

- kontrola funkčnosti a validity,
- technická podpora – tým MUNI-ČD,
- postupný data mining – plnění cílů projektů,
- informační servis stakeholderům.



Struktura dat

typ základní tabulky / sloupce	hrana_z	hrana_do	signatura	hrana_z_cas	hrana_z_datum	hrana_do_cas	hrana_do_datum	z	do	z_cas	z_datum	do_cas	do_datum	opakovani_cesty	dopravni_mod
hrany	•	•	•	•	•	•	•	○	○	○	○	○	○	○	○
hrany-opakovani	•	•	•	•	•	•	•	○	○	○	○	○	○	•	○
hrany-z_do	•	•	•	○	○	○	○	•	•	•	•	•	•	○	○
hrany-z_do-opakovani	•	•	•	○	○	○	○	•	•	•	•	•	•	•	○
signatury	•	•	•	•	•	•	•	○	○	○	○	○	○	○	○
signatury-opakovani	•	•	•	•	•	•	•	○	○	○	○	○	○	•	○
signatury-z_do	•	•	•	○	○	○	○	•	•	•	•	•	•	○	○
signatury-z_do-opakovani	•	•	•	○	○	○	○	•	•	•	•	•	•	•	○
hrany-mod	•	•	•	•	•	•	•	○	○	○	○	○	○	○	•
hrany-opakovani-mod	•	•	•	•	•	•	•	○	○	○	○	○	○	•	•
hrany-z_do-mod	•	•	•	○	○	○	○	•	•	•	•	•	•	○	•
hrany-z_do-opakovani-mod	•	•	•	○	○	○	○	•	•	•	•	•	•	•	•
signatury-mod	•	•	•	•	•	•	•	○	○	○	○	○	○	○	•
signatury-opakovani-mod	•	•	•	•	•	•	•	○	○	○	○	○	○	•	•
signatury-z_do-mod	•	•	•	○	○	○	○	•	•	•	•	•	•	○	•
signatury-z_do-opakovani-mod	•	•	•	○	○	○	○	•	•	•	•	•	•	•	•
signatury-z_do-opakovani-mod	•	•	•	○	○	○	○	•	•	•	•	•	•	•	•

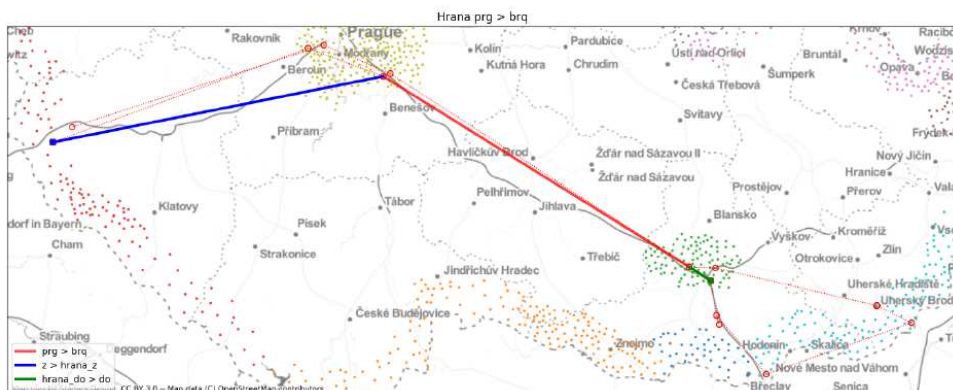
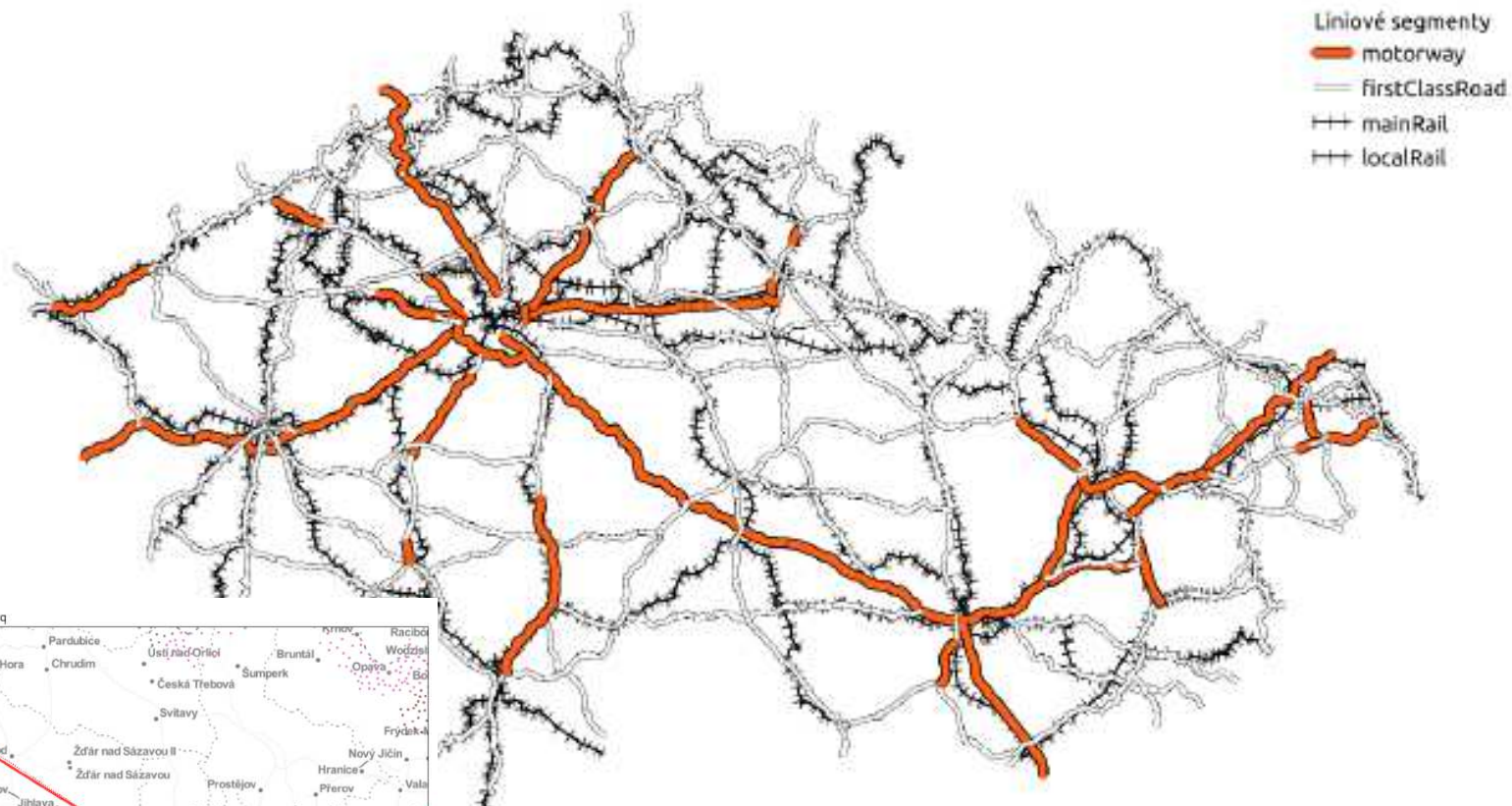
➤ 2 základní náhledy:

- hrany
- signatury

➤ s a bez dopravní módu

Rozlišení módu:

- dálnice,
- silnice 1.tř.,
- železniční koridory,
- hlavní železniční tratě.



n	kritérium / ukazatel	typ_kritéria	úroveň_podrobnosti				jednotka	počet_relevantních_souborů	v_%
1	gdpr_ratio_cestujici	1						1 257 988	100,00%
2	gdpr_ratio_cesty	1						1 257 988	100,00%
3	gdpr_ratio_records	1						1 257 988	100,00%
4	index	1						1 257 988	100,00%
5	mode	1						1 257 988	100,00%
6	name	1						1 257 988	100,00%
7	output_fname	1						1 257 988	100,00%
8	col_cestovali_nkrat	1						193 540	15,38%
9	col_do	4	kraj	okres	orp	pou		1 216 528	96,70%
10	col_do_cas	3	celyden	hodina	pulden			1 216 524	96,70%
11	col_do_datum	4	celeobdobi	datum	denvtydnu	tyden		1 216 528	96,70%
12	col_doba_cesty_hrana_z_do	1					30min	4 612	0,37%
13	col_doba_cesty_z_do	1					30min	147 460	11,72%
14	col_hrana_do	1						1 257 988	100,00%
15	col_hrana_do_cas	3	celyden	hodina	pulden			41 484	3,30%
16	col_hrana_do_datum	4	celeobdobi	datum	denvtydnu	tyden		41 488	3,30%
17	col_hrana_z	1						1 257 988	100,00%
18	col_hrana_z_cas	3	celyden	hodina	pulden			41 484	3,30%
19	col_hrana_z_datum	4	celeobdobi	datum	denvtydnu	tyden		41 488	3,30%
20	col_mod_dopravy	1						628 996	50,00%
21	col_opakovani_cesty	1						532 228	42,31%
22	col_pocet_cest	1						1 064 452	84,62%
23	col_pocet_cestujicich	1						1 257 988	100,00%
24	col_signatura	1						1 257 988	100,00%
25	col_vse	1						78 340	6,23%
26	col_vstup_vystup_cr	1						78 340	6,23%
27	col_vztah_do	1						73 732	5,86%
28	col_vztah_z	1						73 732	5,86%
29	col_vztah_z_do	1						405 508	32,23%
30	col_z	4	kraj	okres	orp	pou		1 216 528	96,70%
31	col_z_cas	3	celyden	hodina	pulden			1 216 524	96,70%
32	col_z_datum	4	celeobdobi	datum	denvtydnu	tyden		1 216 528	96,70%
33	col_zeme_puvodu	3	eu	kontinent	stat			573 708	45,61%
34	col_zpatecni_cesta_hrana_z_do	1						4 612	0,37%
35	col_zpatecni_cesta_z_do	1						73 732	5,86%

➤ 35 různých údajů o datech:

- 32 popisných ukazatelů,
- do 4 úrovní podrobnosti.

Práce se soubory

- vlastní webové rozhraní:
 - výběr formátu,
 - výběr období,
 - filtrování,
 - výběr řazení dle GDPR ratio,
 - náhledy,
 - stahování.

ITREGEP Odlášení ze systému

Order by: gdpr_ratio_ Filetype: xlsx Season: muni4 Basic sample type: hrany Detailed sample type: vse

Criterion 1: col_hrana_ Criterion 1 filter: hodina Criterion 2: none Criterion 2 filter: none

Buttons: Apply filter, Download table in XLS, Download table in CSV, Load selected file

Search:

Path	GDPR cestující	GDPR cesty	GDPR records
./output/muni4/munxlsx/hrany/vse/11d815e3686091	1	1	1
./output/muni4/munxlsx/hrany/vse/14e74eea6b8c	1	1	0.954
./output/muni4/munxlsx/hrany/vse/1502c7b6ba6a34	1	1	0.996
./output/muni4/munxlsx/hrany/vse/1541bb6591c72f	1	1	0.967
./output/muni4/munxlsx/hrany/vse/17729173025ad9	1	1	1
./output/muni4/munxlsx/hrany/vse/211efeb376a59c	1	1	1
./output/muni4/munxlsx/hrany/vse/220238ebb4bc72	1	1	0.952
./output/muni4/munxlsx/hrany/vse/2b3c55eb6523ec	1	1	0.996
./output/muni4/munxlsx/hrany/vse/598aa04fcc0e0	1	1	0.996
./output/muni4/munxlsx/hrany/vse/6430c9db2cf5ab	1	1	0.996

Showing 1 to 11 of 47 entries

Verifikace - formální

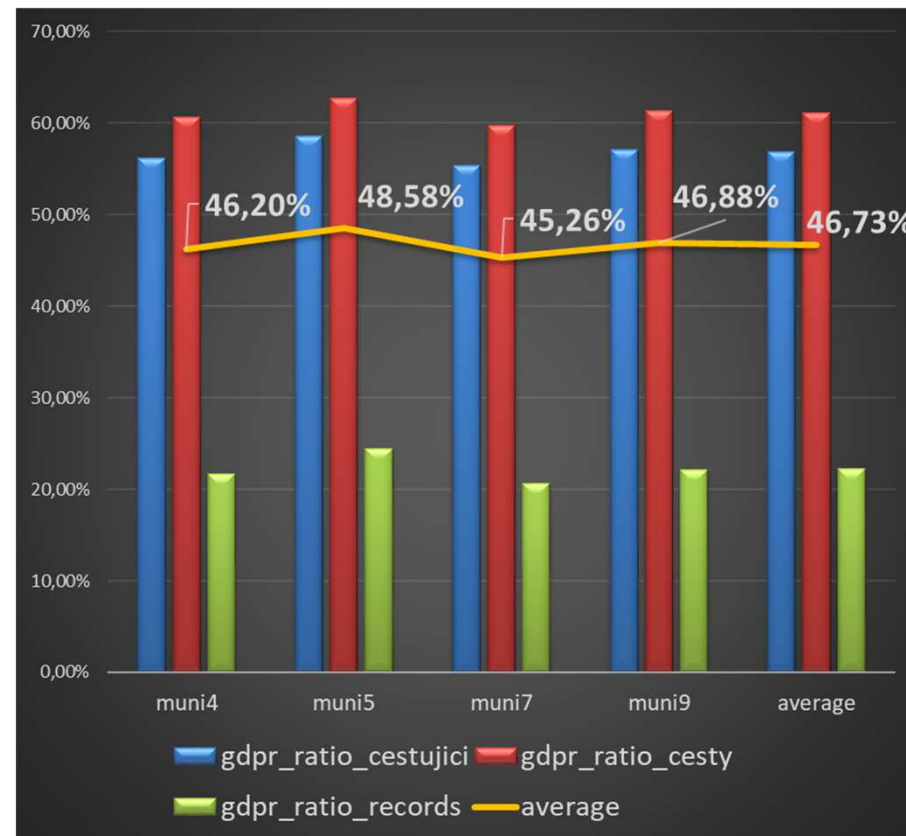
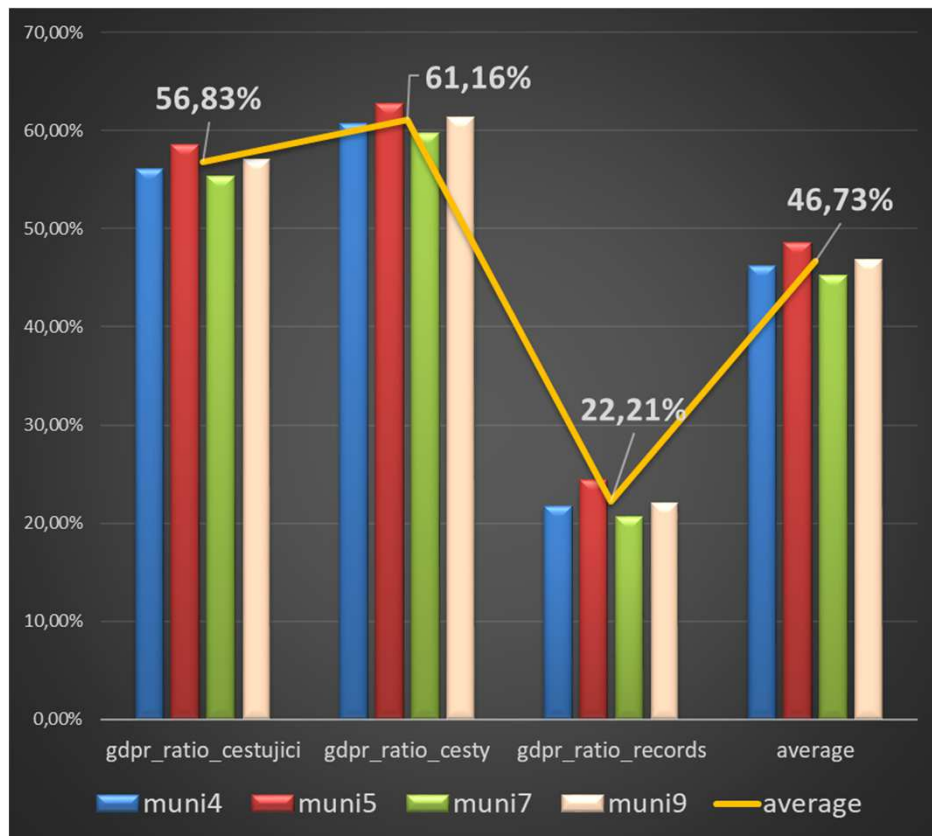
- počty souborů dle jednotlivých ukazatelů v jednotlivých obdobích,
- kontrola odpovídajících datových formátů,
- kontrola absence souborů oproti „mapping_tables“,
- kontrola „GDPR ratio“ pro cesty / cestující / záznamy – informace je nutná u všech souborů.

Formální kontrola – hrany vs. signatury

Dataset - typ	hrany	signatury	celkem
cestovali_nkrat	4 896	4 896	9 792
doba_cesty_hrana_z_do[30min]	576	576	1 152
doba_cesty_z_do[30min]	9 216	9 216	18 432
doba_cesty_z_do[30min]+vztah_z_do	9 216	9 216	18 432
vse	9 792	9 792	19 584
vstup_vystup_cr	9 792	9 792	19 584
vztah_do	9 216	9 216	18 432
vztah_z	9 216	9 216	18 432
vztah_z_do	9 216	9 216	18 432
vztah_z_do+cestovali_nkrat	4 608	4 608	9 216
zeme_puvodu[eu]	9 792	9 792	19 584
zeme_puvodu[eu]+cestovali_nkrat	4 896	4 896	9 792
zeme_puvodu[eu]+vztah_z_do	9 216	9 216	18 432
zeme_puvodu[kontinent]	9 792	9 792	19 584
zeme_puvodu[kontinent]+cestovali_nkrat	4 896	4 896	9 792
zeme_puvodu[kontinent]+vztah_z_do	9 216	9 216	18 432
zeme_puvodu[stat]	9 792	9 792	19 584
zeme_puvodu[stat]+cestovali_nkrat	4 896	4 896	9 792
zeme_puvodu[stat]+vztah_z_do	9 216	9 216	18 432
zpatecni_cesta_hrana_z_do	576	576	1 152
zpatecni_cesta_z_do	9 216	9 216	18 432
celkem	157 248	157 248	314 496

Dataset - náhled	hrany	signatury	celkem
hrany	1 584		1 584
hrany-mod	1 584		1 584
hrany-opakovani	1 008		1 008
hrany-opakovani-mod	1 008		1 008
hrany-z_do	43 776		43 776
hrany-z_do-mod	43 776		43 776
hrany-z_do-opakovani	32 256		32 256
hrany-z_do-opakovani-mod	32 256		32 256
signatury		1 584	1 584
signatury-mod		1 584	1 584
signatury-opakovani		1 008	1 008
signatury-opakovani-mod		1 008	1 008
signatury-z_do		43 776	43 776
signatury-z_do-mod		43 776	43 776
signatury-z_do-opakovani		32 256	32 256
signatury-z_do-opakovani-mod		32 256	32 256
celkem	157 248	157 248	314 496

Formální kontrola – průměrné „GDPR ratio“

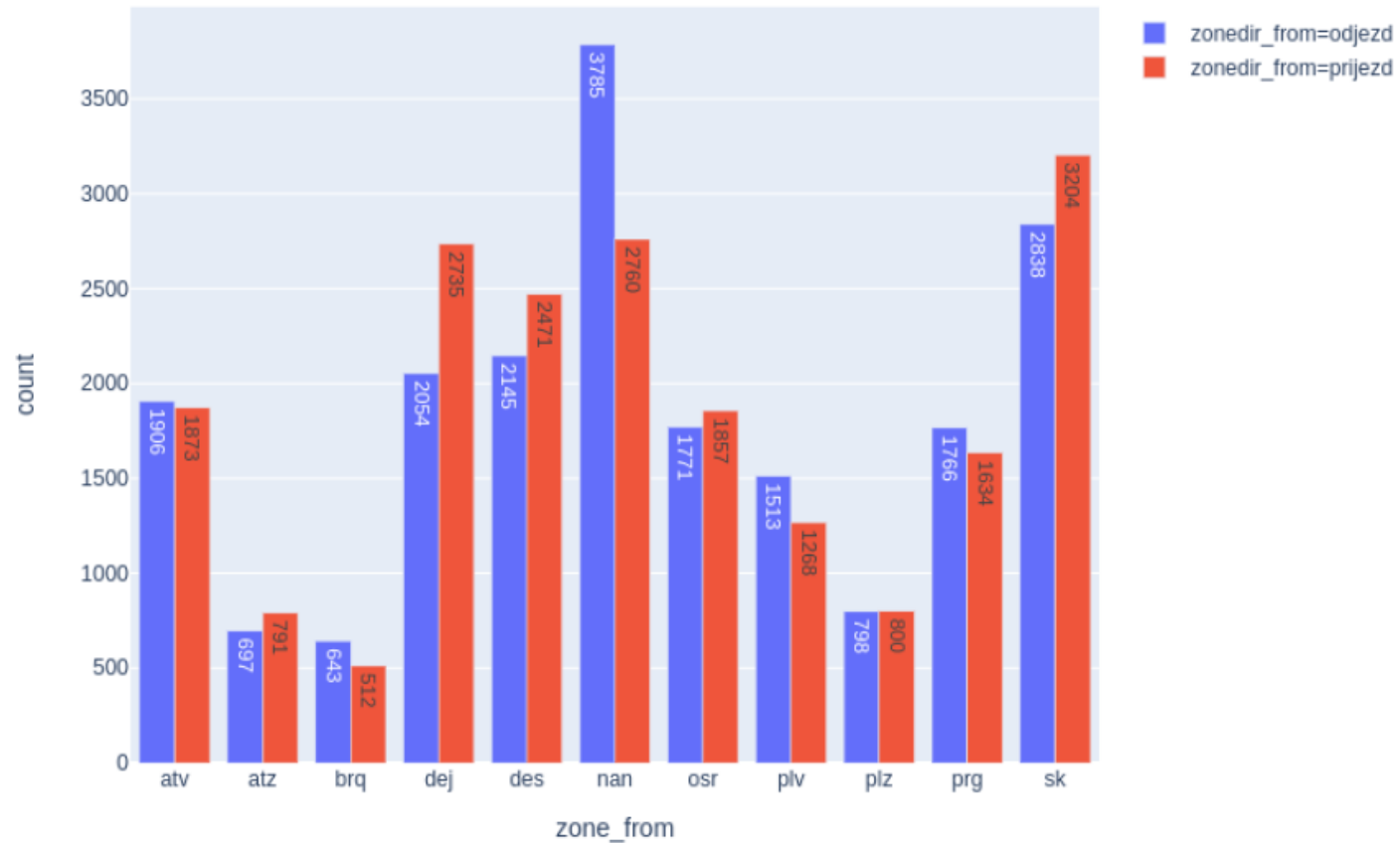


critterion	muni4	muni5	muni7	muni9	average
gdpr_ratio_cestujici	56,17%	58,60%	55,39%	57,16%	56,83%
gdpr_ratio_cesty	60,72%	62,78%	59,73%	61,41%	61,16%
gdpr_ratio_records	21,71%	24,38%	20,67%	22,08%	22,21%
average	46,20%	48,58%	45,26%	46,88%	46,73%

Verifikace - věcná

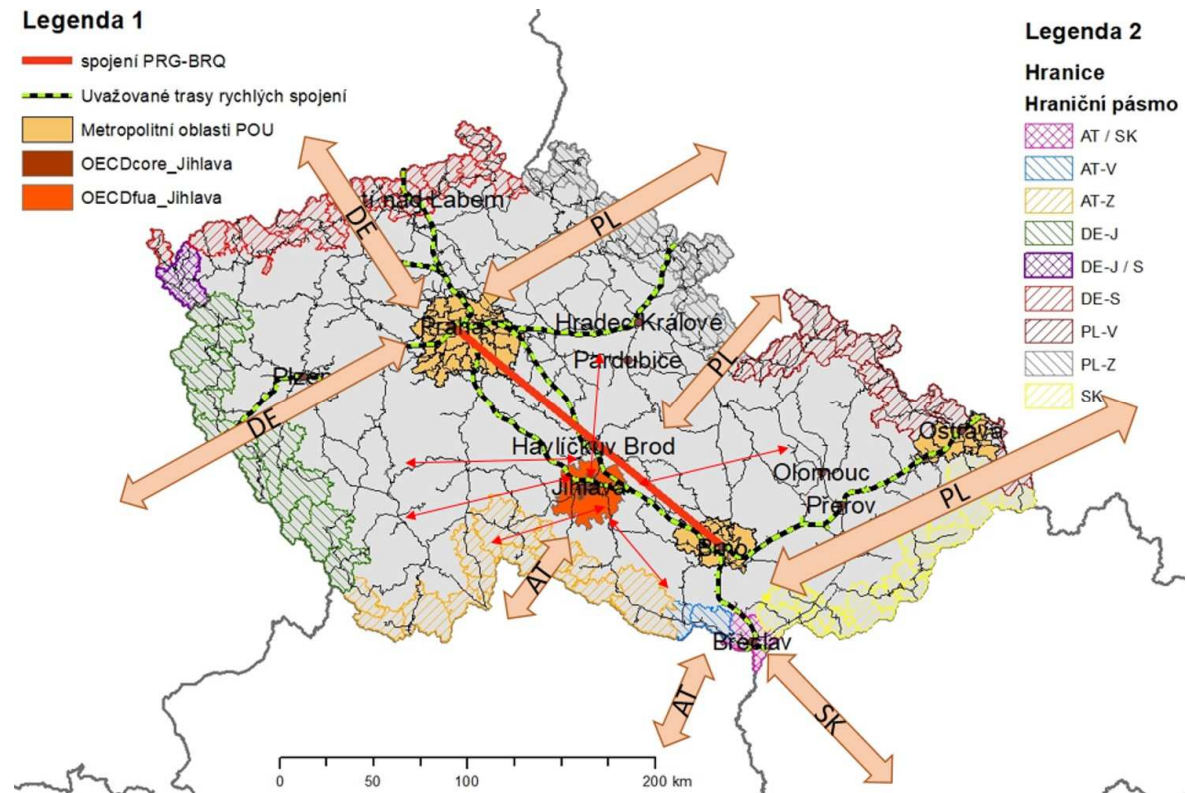
- procentuální výskyt n/a v ukazateli „mode“,
- kontrola součtů cest / cestujících / záznamů napříč obdobími,
- kontrola minimálních, maximálních hodnot v jednotlivých datasetech,
- validita souborů dle „GDPR ratio“,
- kontrola mocnosti směrů a protisměrů napříč časovými obdobími,
- kontrola jízdní doby,
- kontrola dopravních špiček dle průběhu dne/týdne,
- kontrola nejvýznamnějších Origin-Destination,
- kontrola zastoupení „původu SIM“ na hranách,
- kontrola salda dojížděky a vyjížděky.

Věcná kontrola – saldo dojížděky a vyjížděky ze zahraničí



Opce: Big data - Jihlava

- rozšíření sledování – terminál VRT Jihlava,
- zachování kompatibilních parametrů datasetů.



3. Small data

- specifická data ze spotřebitelských šetření
- veřejná soutěž:
 - tržní konzultace,
 - oslovení vybraných dodavatelů,
 - soutěž o cenu dodávky.

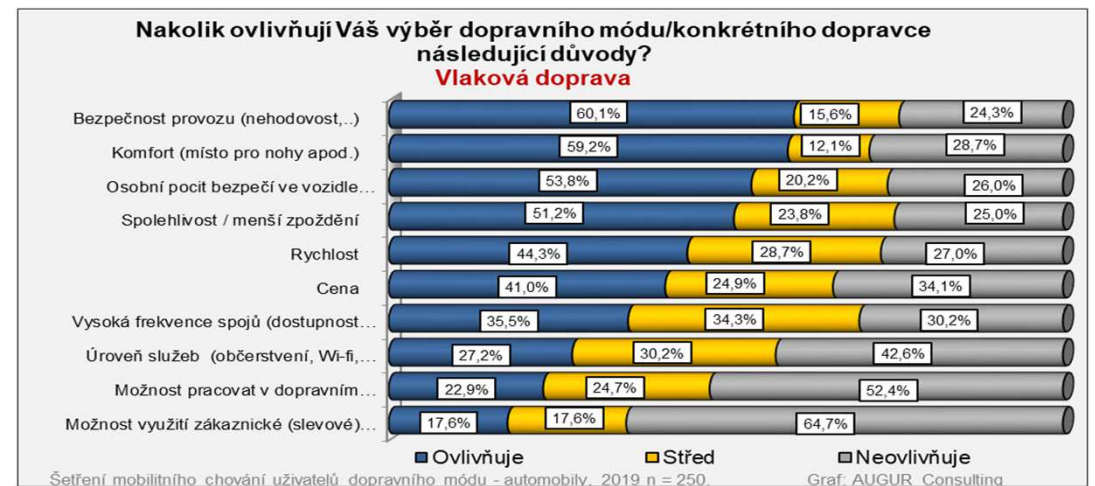
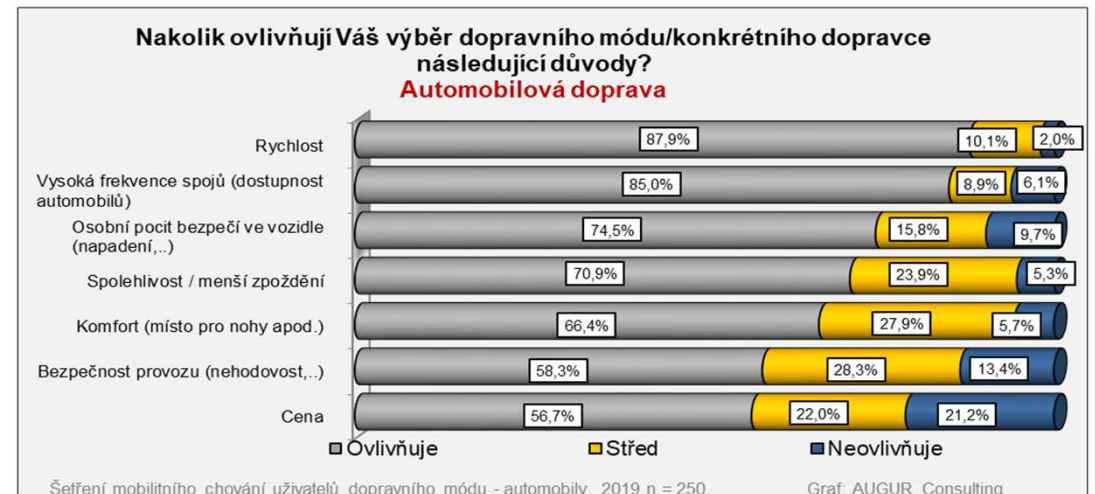
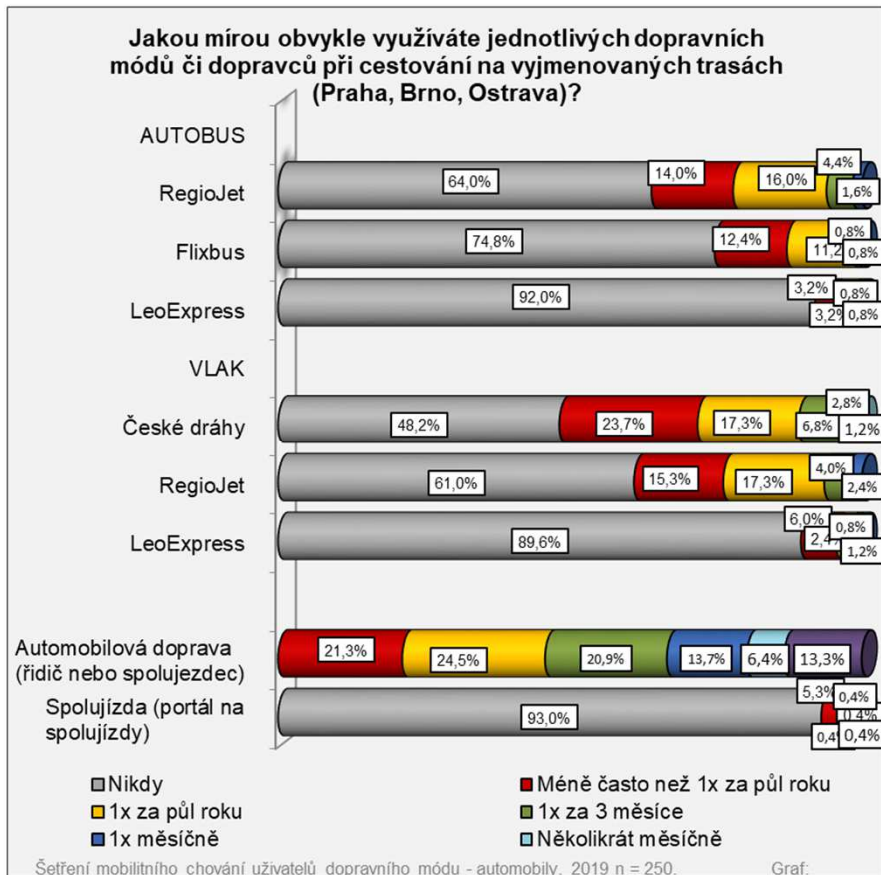
- dodavatel:



- dodávka: srpen – prosinec 2019

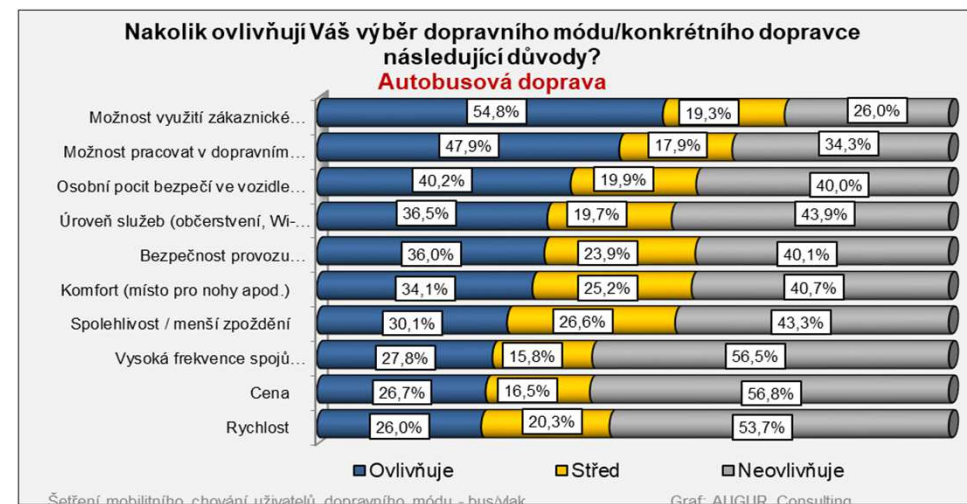
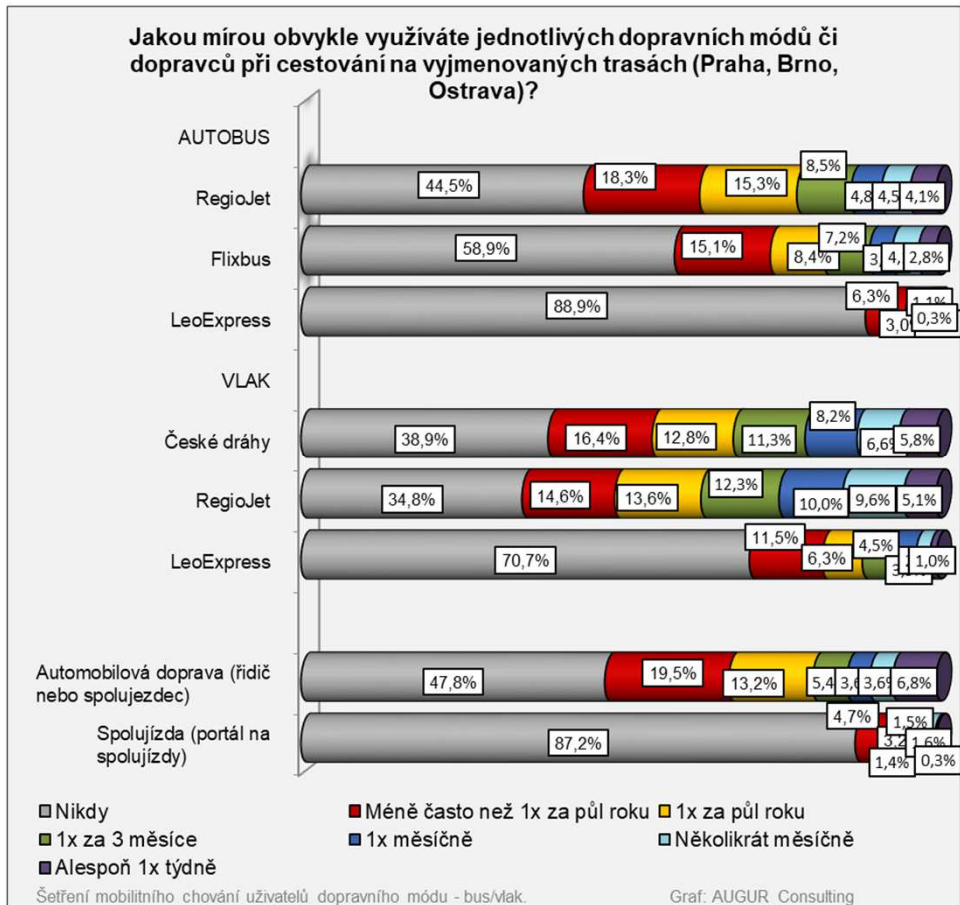
Šetření č. 1 - mobilitní chování uživatelů dopravního módu – automobily

- **Cíl:** faktory podmiňující využití konkrétního dopravního prostředku
- **Respondenti:** 250 (systematický výběr)
- **Metoda:** face-to-face standardizovaný dotazník
- **Místo:** D1, D46, D35 + metropole
- **Zjištění:**
 - socioekonomická identifikace,
 - charakteristika aktuální cesty,
 - mobilitní chování a rozhodování.



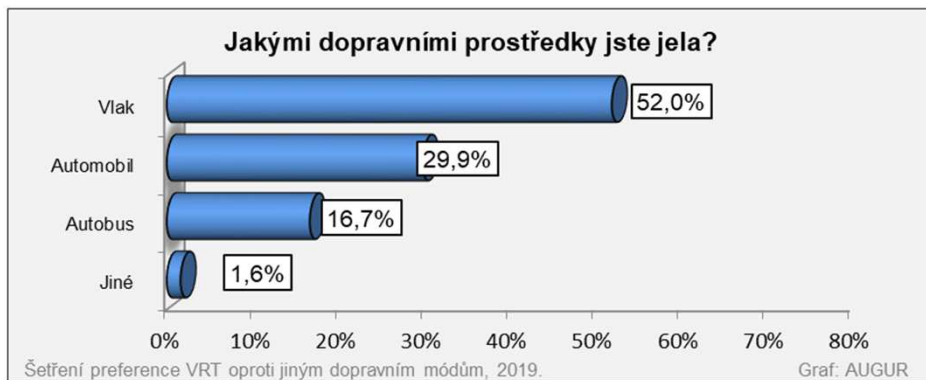
Šetření č. 2 - mobilitní chování uživatelů dopravního módu – bus/vlak

- **Cíl:** faktory podmiňující využití konkrétního dopravního prostředku
- **Respondenti:** 1 000 (systematický výběr)
- **Metoda:** face-to-face standardizovaný dotazník
- **Místo:** nástupní stanice Praha, Brno, Ostrava
- **Zjištění:**
 - socioekonomická identifikace,
 - charakteristika aktuální cesty,
 - mobilitní chování a rozhodování.



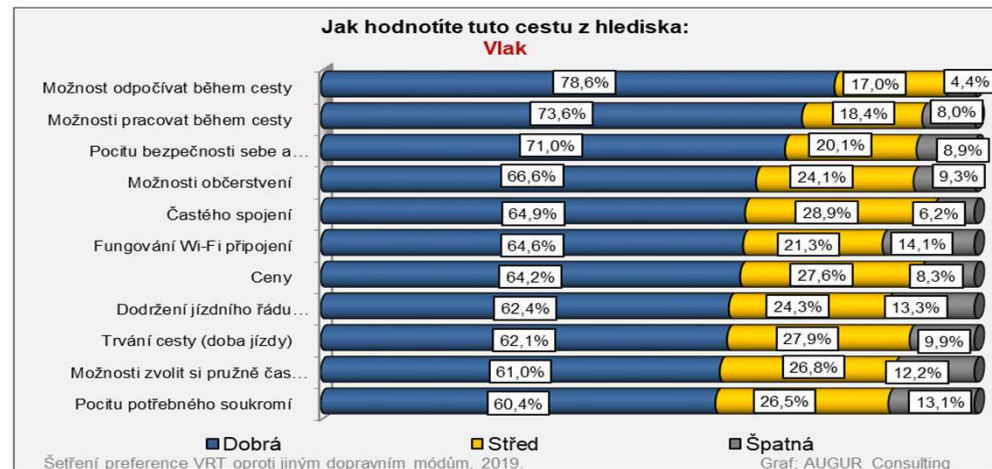
Šetření č. 3 - preference VRT oproti jiným dopravním módům

- **Cíl:** volba VRT oproti IAD, bus, letadlo a konvenční vlak
- **Respondenti:** 1 200 (kvótní výběr)
- **Metoda:** face-to-face standardizovaný dotazník
- **Místo:** Praha, Brno, Ostrava
- **Zjištění:**
 - socioekonomická identifikace,
 - charakteristika aktuální cesty,
 - potenciál změny mobilního chování.



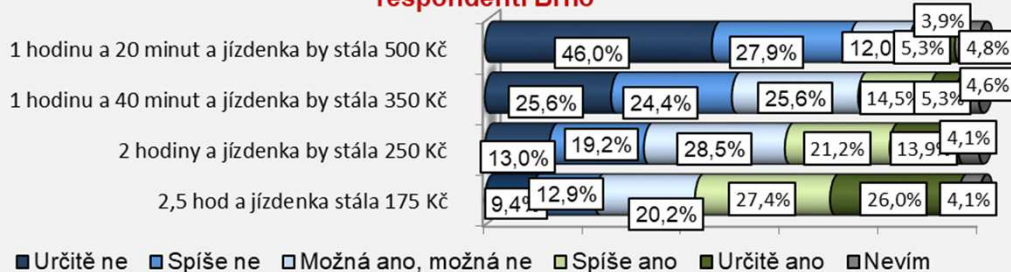
Jak dlouho Vám trvala cesta od dveří ke dveřím?

		Průměrně minut
Trasa	Brno-Praha	216,9
	Brno-Ostrava	197,0
	Praha-Brno	226,6
	Praha-Ostrava	264,2
	Ostrava-Brno	205,7
	Ostrava-Praha	277,8



S jakou pravděpodobností byste pro svoji cestu do Prahy zvolil/a vlak, kdyby cesta do Prahy trvala...

respondenti Brno



Šetření preference VRT oproti jiným dopravním módům, 2019.

Graf: AUGUR Consulting

S jakou pravděpodobností byste pro svoji cestu do Prahy zvolil/a vlak, kdyby cesta do Prahy trvala...

respondenti Ostrava

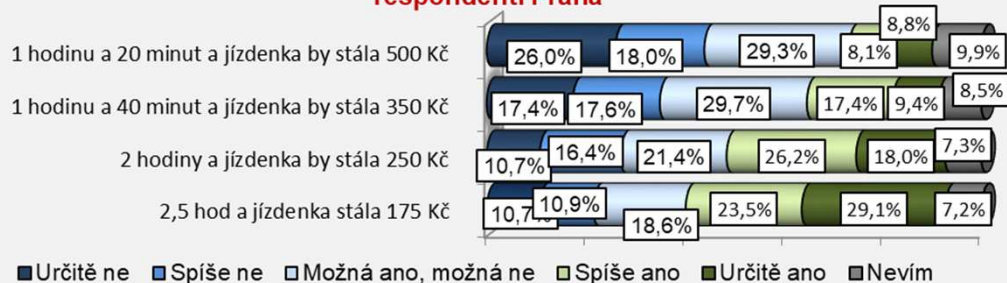


Šetření preference VRT oproti jiným dopravním módům, 2019.

Graf: AUGUR Consulting

S jakou pravděpodobností byste pro svoji cestu do Brna zvolil/a vlak, kdyby cesta do Brna trvala...

respondenti Praha

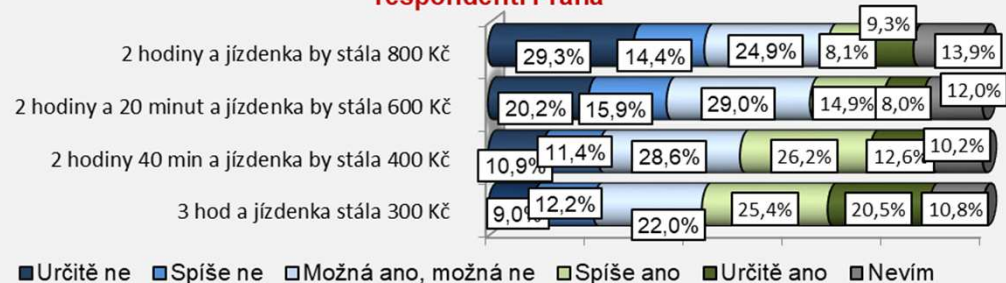


Šetření preference VRT oproti jiným dopravním módům, 2019.

Graf: AUGUR Consulting

S jakou pravděpodobností byste pro svoji cestu do Ostravy zvolil/a vlak, kdyby cesta do Ostravy trvala...

respondenti Praha



Šetření preference VRT oproti jiným dopravním módům, 2019.

Graf: AUGUR Consulting

Šetření č. 4 - mobilitní chování domácích a zahraničních návštěvníků

- **Cíl:** vliv dostupnosti destinace na volbu místa návštěvy a cestovní zvyklosti návštěvníků
- **Respondenti:** 1 800 (kvótní výběr)
- **Metoda:** face-to-face standardizovaný dotazník
- **Místo:** turistické destinace v oblasti Prahy, Brna, Ostravy
- **Zjištění:**
 - socioekonomická identifikace,
 - charakteristika aktuální cesty,
 - současné rozhodování,
 - význam dostupnosti a její změny (VRT).

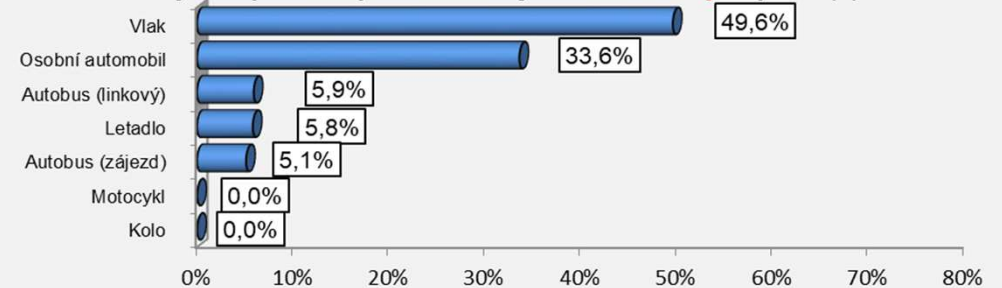
Jaký je hlavní účel Vaší návštěvy **Prahy**?



Šetření mobilního chování domácích a zahraničních návštěvníků, 2019.

Graf: AUGUR

Jakým dopravním prostředkem jste se do **Prahy** dopravil(a)?



Šetření mobilního chování domácích a zahraničních návštěvníků, 2019.

Graf: AUGUR Consulting

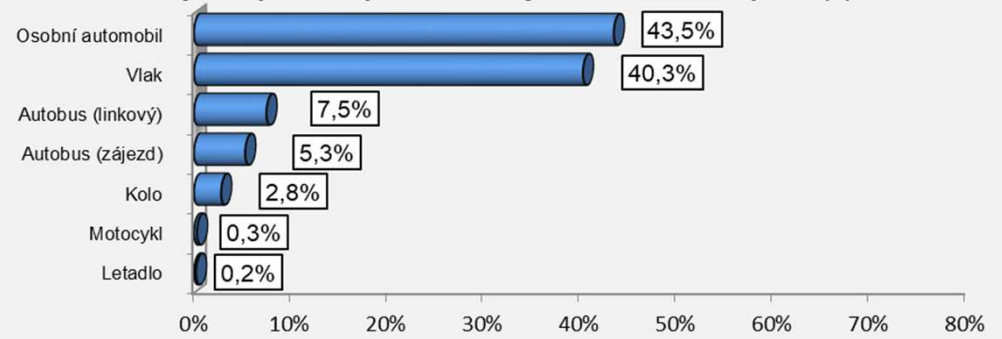
Jaký je hlavní účel Vaší návštěvy **Brna**?



Šetření mobilního chování domácích a zahraničních návštěvníků, 2019.

Graf: AUGUR

Jakým dopravním prostředkem jste se do **Brna** dopravil(a)?



Šetření mobilního chování domácích a zahraničních návštěvníků, 2019.

Graf: AUGUR Consulting

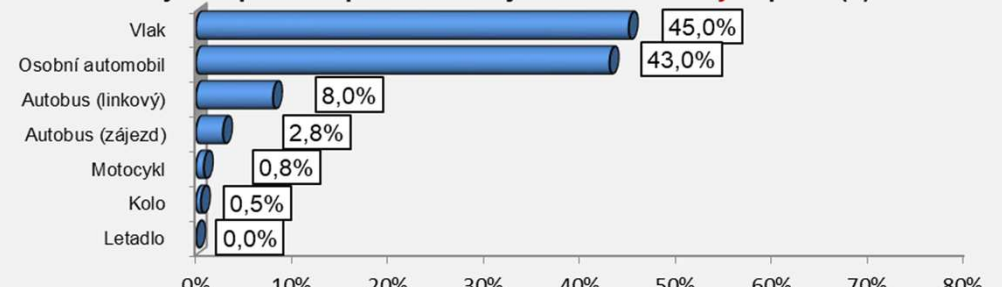
Jaký je hlavní účel Vaší návštěvy **Ostravy**?



Šetření mobilního chování domácích a zahraničních návštěvníků, 2019.

Graf: AUGUR

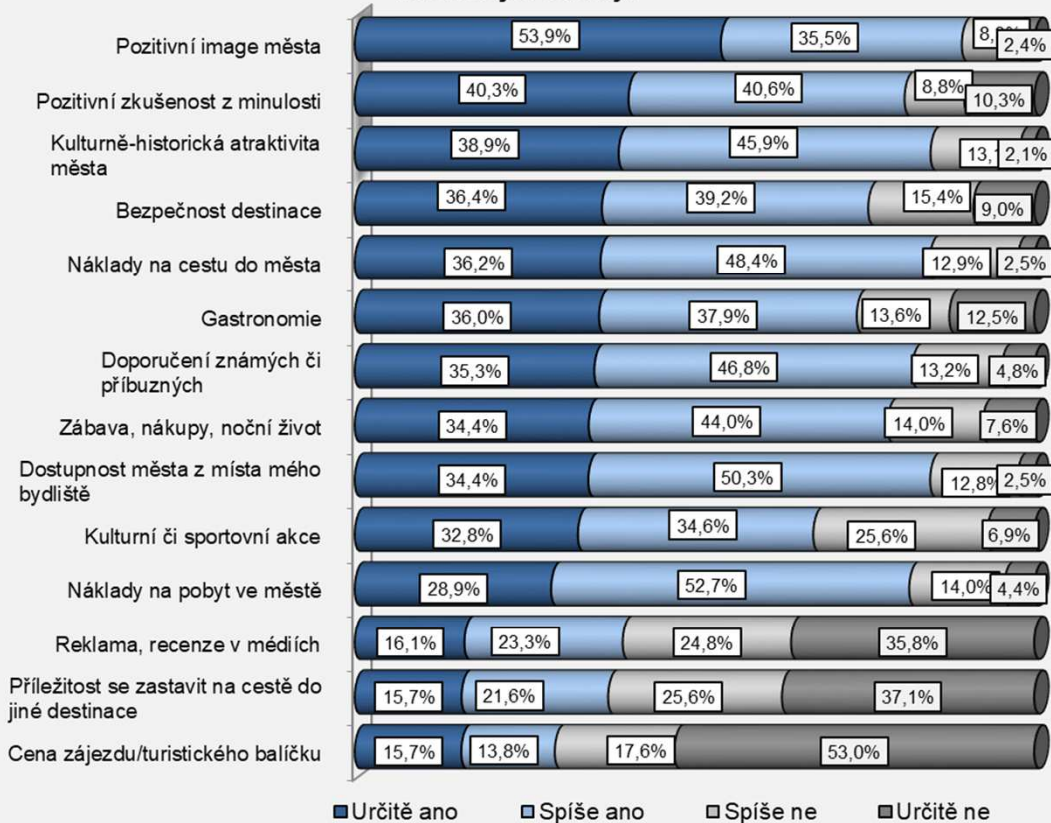
Jakým dopravním prostředkem jste se do **Ostravy** dopravil(a)?



Šetření mobilního chování domácích a zahraničních návštěvníků, 2019.

Graf: AUGUR Consulting

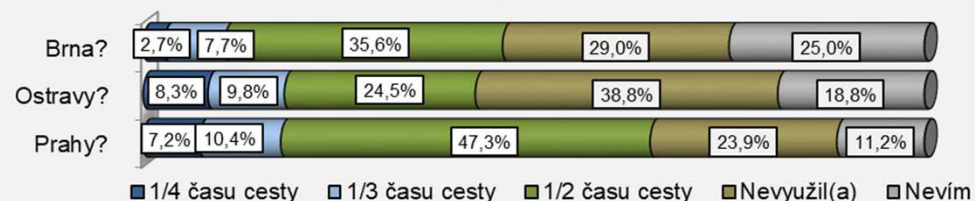
Do jaké míry Vás při rozhodování o cestě do Prahy ovlivnily následující faktory?



Šetření mobilního chování domácích a zahraničních návštěvníků, 2019.

Graf: AUGUR Consulting

Při jaké hypotetické úspoře času byste již využil(a) rychlovlak na cestu do..



Šetření mobilního chování domácích a zahraničních návštěvníků, 2019.

Graf: AUGUR Consulting

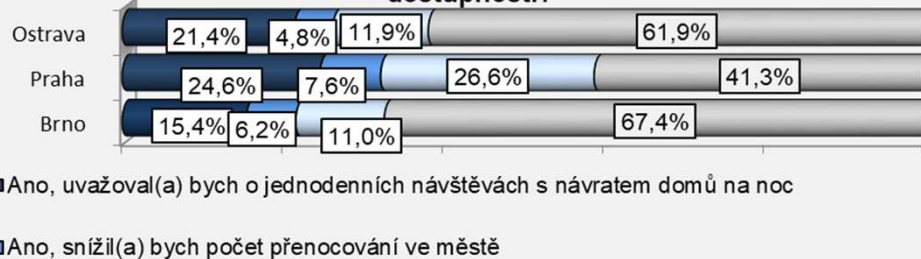
Zvýšíte v takovém případě Vaši frekvenci cest do této destinace za rekreaci, cestovním ruchem, za prací či pro účely tranzitu?



Šetření mobilního chování domácích a zahraničních návštěvníků, 2019.

Graf: AUGUR Consulting

Změnila by se délka Vašeho pobytu ve městě, při zlepšení časové dostupnosti?



Šetření mobilního chování domácích a zahraničních návštěvníků, 2019.

Graf: AUGUR Consulting

Šetření č. 5 - mobilitní chování účastníků cestovního ruchu – Jižní Korea

- **Cíl:** cestovní zvyklosti v ČR a střední Evropě + vliv dostupnosti destinace na rozhodování
- **Respondenti:** 20 (rekrutace – jihokorejští turisté)
- **Metoda:** focus group
- **Místo:** Praha, Brno
- **Zjištění:**
 - faktory volby cílové destinace,
 - charakter mobilitního chování v destinaci,
 - role módu dopravy při transferech.

Šetření č. 6 - preference VRT oproti letadlu

- **Cíl:** faktory podmiňující substituce VRT s letadlem a autobusem
- **Respondenti:** 30 (rekrutace)
- **Metoda:** focus group
- **Místo:** Praha - cestující na lince Praha–Mnichov
- **Zjištění:**
 - charakteristika aktuální cesty,
 - mobilní chování a rozhodování,
 - potenciál preference VRT.

Opce: Šetření č. 7 - změny mobilityního a rezidenčního chování

- **Cíl:** míra vlivu zlepšené nabídky železniční dopravy na změnu mobilityního a rezidenčního chování
- **Respondenti:** 400 (systematický výběr)
- **Metoda:** face-to-face standardizovaný dotazník
- **Místo:** železniční linka Plzeň–Praha
- **Zjištění:**
 - socioekonomická identifikace
 - mobilityní chování v minulosti a nyní
 - motivy/důvody mobilityní/rezidenční změny

Opce: Šetření č. 8 - potenciální změny mobilityního chování na RS1

- **Cíl:** postoje cestujících k různým dopravním modelům VRT ve vazbě na jejich současné dopravní chování
- **Respondenti:** 1 200 (systematický výběr)
- **Metoda:** face-to-face standardizovaný dotazník
- **Místo:** D1 + nástupní stanice Praha, Brno (vlak, bus)
- **Zjištění:**
 - mobilityní chování a rozhodování,
 - preference dopravního modelu VRT,
 - míra intermodálního přesunu.

Opce: Šetření č. 9 - faktory potenciální změny mobilityního chování na RS1

- **Cíl:** postoje cestujících k různým dopravním modelům VRT ve vazbě na jejich současné dopravní chování
- **Respondenti:** 50 (rekrutace – cestující mezi Prahou a Brnem)
- **Metoda:** focus group
- **Místo:** Praha, Brno
- **Zjištění:**
 - mobilityní chování a rozhodování,
 - preference dopravního modelu VRT,
 - míra intermodálního přesunu.

Děkuji za pozornost!

Nová mobilita – vysokorychlostní dopravní systémy a dopravní chování populace

Operační program Výzkum, vývoj a vzdělávání, výzva č. 26: Dlouhodobá mezisektorová spolupráce
CZ.02.1.01/0.0/0.0/16 026/0008430



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

