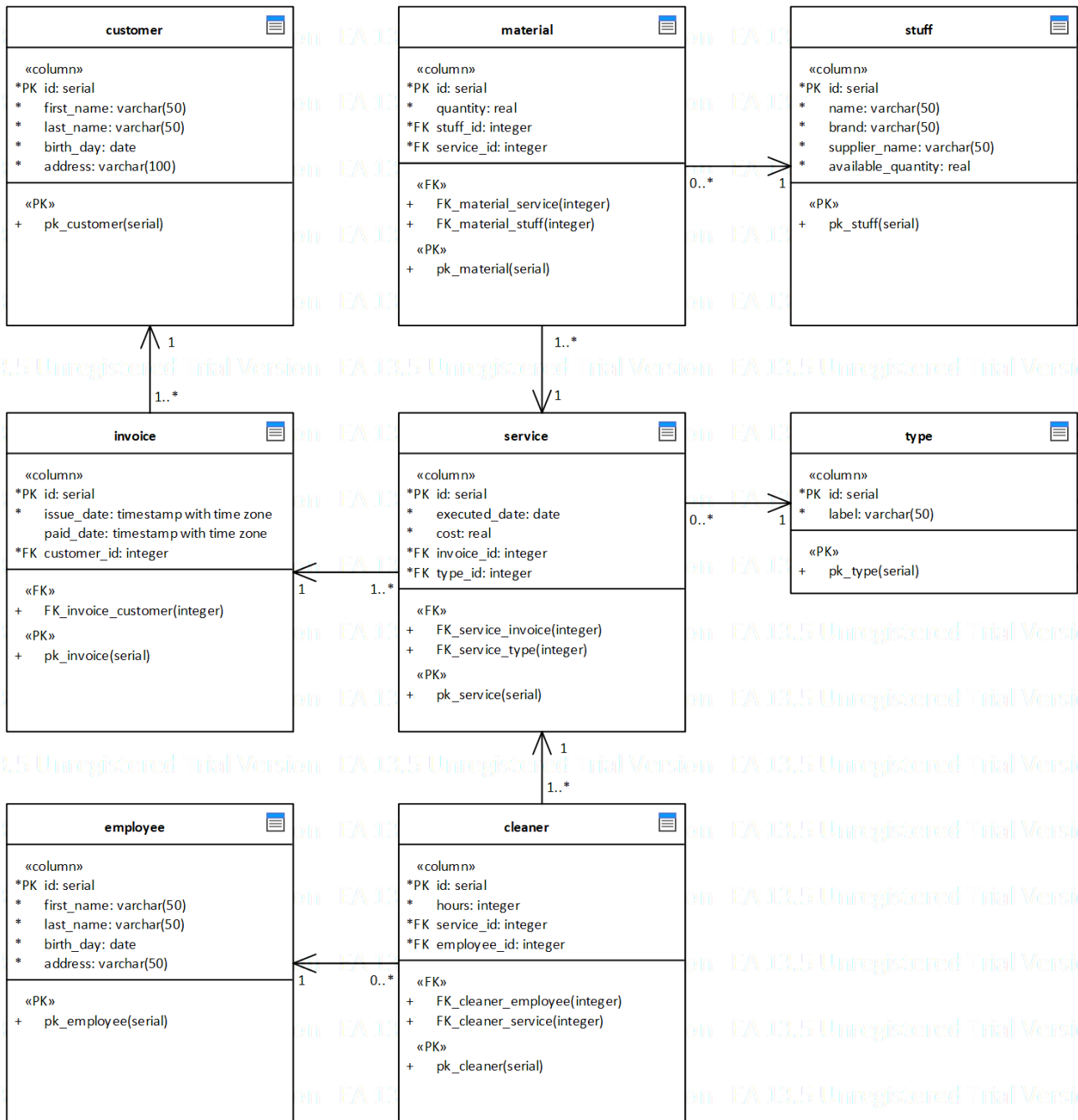


Analytický report

Ondrej Harnúšek

Physical model



Spôsob generovania dát

Dáta boli generované prostredníctvom programu Datanamic. Nominálne atribúty boli generované náhodne z preddefinovaných zoznamov, s dôrazom na čo najväčšiu podobnosť danej doméne. Numerické atribúty boli generované náhodne až na atribúty cudzích kľúčov. Tie boli generované sekvenčne z dôvodu potreby prepojiť všetky záznamy.

invoice

count: 1 000 000

```
SELECT count(*) FROM invoice
```

COLUMN NAME	DATATYPE	COMMENTS
 id	serial	<p>min : 1 <pre>SELECT min(id) FROM invoice</pre></p> <p>max : 1000000 <pre>SELECT max(id) FROM invoice</pre></p>
 issue_date	timestamp with time zone	<p>min : 1990 <pre>SELECT min(EXTRACT(YEAR FROM issue_date)) FROM invoice</pre></p> <p>max : 2017 <pre>SELECT max(EXTRACT(YEAR FROM issue_date)) FROM invoice</pre></p> <p>avg : 2003.30886 <pre>SELECT avg(EXTRACT(YEAR FROM issue_date)) FROM invoice</pre></p> <p>dolný kvartil : 1996 <pre>SELECT percentile_cont(0.25) within group (order by EXTRACT(YEAR FROM issue_date)) FROM invoice</pre></p> <p>medián : 2003 <pre>SELECT percentile_cont(0.5) within group (order by EXTRACT(YEAR FROM issue_date)) FROM invoice</pre></p> <p>horný kvartil : 2010 <pre>SELECT percentile_cont(0.75) within group (order by EXTRACT(YEAR FROM issue_date)) FROM invoice</pre></p>

COLUMN NAME	DATATYPE	COMMENTS
 paid_date	timestamp with time zone	<p>min : 1990 SELECT min(EXTRACT(YEAR FROM paid_date)) FROM invoice</p> <p>max : 2017 SELECT max(EXTRACT(YEAR FROM paid_date)) FROM invoice</p> <p>avg : 2003.2962930824 SELECT avg(EXTRACT(YEAR FROM paid_date)) FROM invoice</p> <p>dolný kvartil : 1996 SELECT percentile_cont(0.25) within group (order by EXTRACT(YEAR FROM paid_date)) FROM invoice</p> <p>medián : 2003 SELECT percentile_cont(0.5) within group (order by EXTRACT(YEAR FROM paid_date)) FROM invoice</p> <p>horný kvartil : 2010 SELECT percentile_cont(0.75) within group (order by EXTRACT(YEAR FROM paid_date)) FROM invoice</p>
 customer_id	integer	<p>min : 1 SELECT min(customer_id) FROM invoice</p> <p>max : 100000 SELECT max(customer_id) FROM invoice</p> <p>avg : 50002.796205 SELECT avg(customer_id) FROM invoice</p> <p>dolný kvartil : 25022 SELECT percentile_cont(0.25) within group (order by customer_id) FROM invoice</p> <p>medián : 49997 SELECT percentile_cont(0.5) within group (order by customer_id) FROM invoice</p> <p>horný kvartil : 74995 SELECT percentile_cont(0.75) within group (order by customer_id) FROM invoice</p>



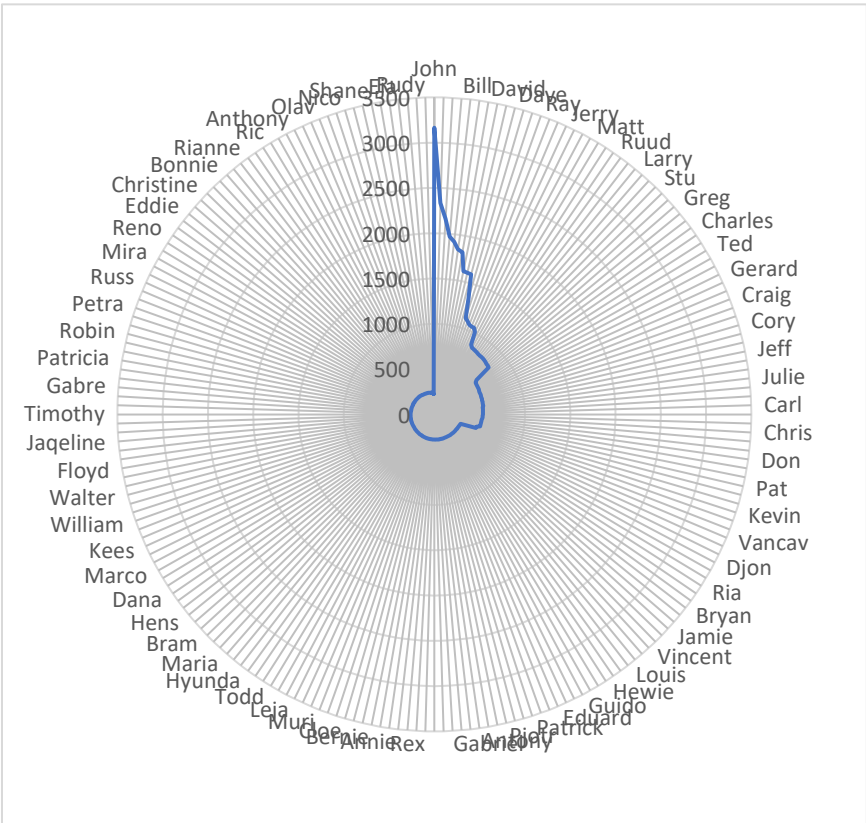

cleaner

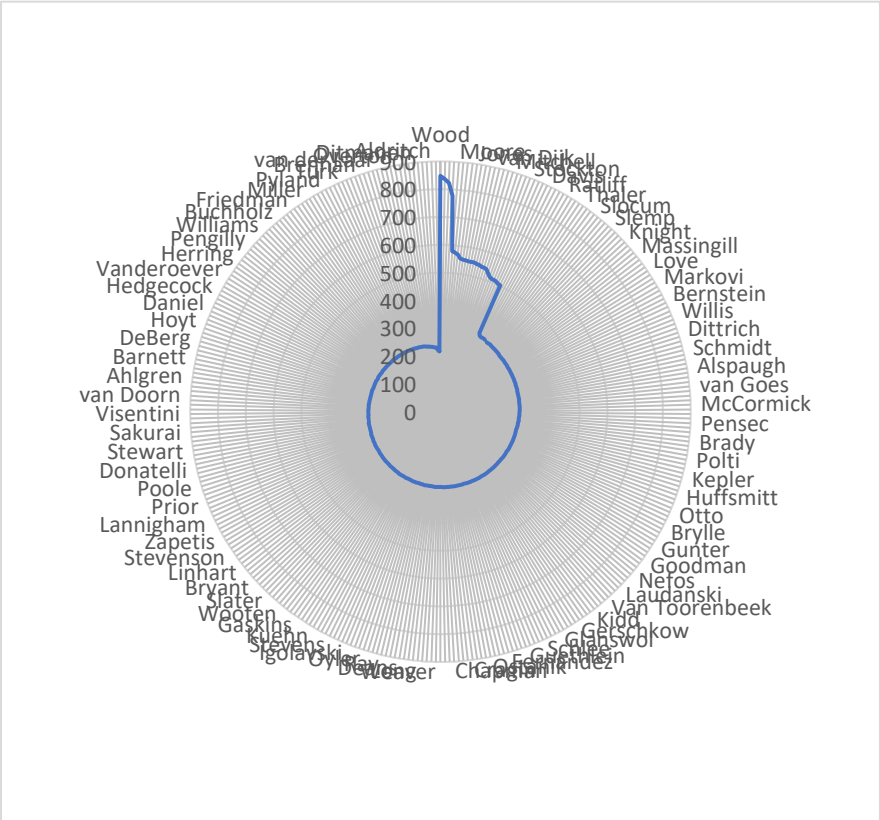


count: 2 000 000

COLUMN NAME	DATATYPE	COMMENTS
 id	serial	min : 1 max : 2 000 000
 hours	integer	min : 2 max : 3600 avg : 1800.835713000000000000 dolný kvartil : 901 medián : 1801 horný kvartil : 2700
 service_id	integer	min : 1 max : 2000000 avg : 1000000.5 dolný kvartil : 500000.75 medián : 1000000.5 horný kvartil : 1500000.25
 employee_id	integer	min : 1 max : 100000 avg : 50006.4812835 dolný kvartil : 24976.75 medián : 50026 horný kvartil : 75013

customer

count: 100 000

COLUMN NAME	DATATYPE	COMMENTS
 id	serial	<p>min : 1</p> <p>max : 100000</p>
 first_name	varchar(50)	<p>the number of repeating values - how many times :</p> <p>1-3165 1-524 1-525 1-526 1-530 1-536 1-539 1-548 1-551 1-552 1-554 1-555 1-559 1-562 1-565 1-566 1-567 1-571 1-573 1-599 1-792 1-799 1-800 1-808 1-813 1-817 1-823 1-829 1-835 1-839 1-844 1-869 1-1014 1-1050 1-1052 1-1062 1-1094 1-1115 1-1339 1-1599 1-1605 1-1612 1-1816 1-1842 1-1918 1-1971 1-2166 1-2334 1-226 1-238 1-246 1-247 1-250 1-274 1-281 1-288 1-289 1-292 1-293 1-294 1-295 1-300 1-301 1-482 1-486 1-520 2-538 2-556 2-243 2-245 2-296 2-298 2-248 2-249 2-546 2-269 2-270 2-273 2-820 2-275 2-305 2-285 2-532 2-549 2-290 2-291 2-537 2-523 2-236 3-267 3-272 3-287 3-276 3-255 3-256 3-268 4-280 4-259 4-277 4-278 4-251 4-264 4-263 4-257 5-282 5-265 5-260 5-261 5-279 5-262 5-283 6-258 7-266 7-271</p> <p>SELECT count(count) the_number_of_repeating_values, count how_many_times FROM (SELECT count(first_name) FROM customer GROUP BY first_name) subtable GROUP BY 2 ORDER BY 1;</p> 
	varchar(50)	the number of repeating values - how many times :



COLUMN NAME	DATATYPE	COMMENTS
last_name		<p>1-217 1-221 1-230 1-234 1-235 1-240 1-241 1-243 1-247 1-251 1-293 1-307 1-316 1-502 1-505 1-509 1-517 1-526 1-538 1-541 1-542 1-546 1-547 1-556 1-568 1-574 1-579 1-778 1-823 1-837 1-848 2-511 2-296 2-232 2-297 2-298 2-552 2-245 2-242 2-304 2-281 2-238 2-236 2-295 3-282 3-305 3-550 3-284 3-248 3-291 3-290 3-244 4-249 4-246 4-287 4-267 4-258 4-264 4-283 4-300 4-254 4-253 5-262 5-266 5-279 5-289 5-292 5-250 5-260 6-274 6-288 6-255 7-286 7-271 7-270 7-252 7-257 8-272 8-278 8-256 8-285 8-265 8-276 8-275 9-269 9-273 9-259 9-261 9-277 10-263 11-280 11-268</p> 
 birth_day	date	<p>min : 1950</p> <p>max : 2000</p> <p>avg : 1974.80065</p> <p>dolný kvartil : 1962</p> <p>medián : 1975</p> <p>horný kvartil : 1987</p>
 address	varchar(100)	<p>the number of repeating values - how many times :</p> <p>1-292 1-293 1-294 1-296 1-297 1-298 1-299 1-302 1-304 1-306 1-308 1-317 1-318 1-320 1-326 1-341 1-363 1-367 1-370 1-372 1-374 1-376 1-379 1-380 1-382 1-383 1-644 2-314 2-328 2-384 2-371 2-388 2-368 2-369 2-325 2-355 2-323 2-313 2-346 2-364 2-366 3-375 3-365 3-321 3-378 3-377 3-315 3-348 3-349 3-312 3-373 3-354 3-356 3-358 3-360 3-309 4-319 4-327 4-322 4-362 4-333 4-329 4-351 4-311 4-331 5-324 5-330 5-336 5-340 5-350 5-353 5-357 5-310 6-316 6-334 6-338 7-335</p>


COLUMN NAME	DATATYPE	COMMENTS
		7-344 7-337 7-342 7-359 7-361 8-332 9-343 9-345 9-347 9-352 12-339

employee

count: 100 000

employee


COLUMN NAME	DATATYPE	COMMENTS
 id	serial	min : 1 max : 100000
 first_name	varchar(50)	the number of repeating values - how many times : 1-3239 1-230 1-233 1-520 1-522 1-523 1-528 1-532 1-536 1-537 1-538 1-542 1-548 1-550 1-552 1-554 1-556 1-565 1-570 1-572 1-579 1-602 1-779 1-783 1-799 1-804 1-806 1-811 1-822 1-856 1-872 1-904 1-1019 1-1038 1-1051 1-1064 1-1084 1-1125 1-1304 1-1561 1-1563 1-1625 1-1850 1-1881 1-1889 1-1898 1-2114 1-2527 1-223 1-243 1-244 1-249 1-250 1-252 1-265 1-277 1-278 1-281 1-285 1-293 1-296 1-303 1-305 1-309 1-310 1-311 1-314 1-318 1-497 1-500 1-506 1-509 1-512 1-513 1-517 1-518 1-519 2-246 2-262 2-812 2-264 2-290 2-824 2-272 2-553 2-511 2-288 2-297 2-286 2-540 2-255 2-256 2-245 2-563 2-235 2-260 3-279 3-237 3-291 3-247 3-257 3-261 3-301 3-270 3-283 3-284 4-271 4-259 4-280 4-258 4-251 4-289 4-253 4-268 4-276 4-267 5-273 5-274 5-254 5-282 5-263 5-287 5-269 1-3239 1-230 1-233 1-520 1-522 1-523 1-528 1-532 1-536 1-537 1-538 1-542 1-548 1-550 1-552 1-554 1-556 1-565 1-570 1-572 1-579 1-602 1-779 1-783 1-799 1-804 1-806 1-811 1-822 1-856 1-872 1-904 1-1019 1-1038 1-1051 1-1064 1-1084 1-1125 1-1304 1-1561 1-1563 1-1625 1-1850 1-1881 1-1889 1-1898 1-2114 1-2527 1-223 1-243 1-244 1-249 1-250 1-252 1-265 1-277 1-278 1-281 1-285 1-293 1-296 1-303 1-305 1-309 1-310 1-311 1-314 1-318 1-497 1-500 1-506 1-509 1-512 1-513 1-517 1-518 1-519 2-246 2-262 2-812 2-264 2-290 2-824 2-272 2-553 2-511 2-288 2-297 2-286 2-540 2-255 2-256 2-245 2-563 2-235 2-260 3-279 3-237 3-291 3-247 3-257 3-261 3-301 3-270 3-283 3-284 4-271 4-259 4-280 4-258 4-251 4-289 4-253 4-268 4-276 4-267 5-273 5-274 5-254 5-282 5-263 5-287 5-269

COLUMN NAME	DATATYPE	COMMENTS
 last_name	varchar(50)	<p>the number of repeating values - how many times :</p> <p>1-322 1-319 1-305 1-301 1-293 1-825 1-218 1-237 1-817 1-816 1-245 1-207 1-247 1-566 1-563 1-557 1-252 1-550 1-548 1-538 1-535 1-533 1-528 1-527 1-847 1-525 1-522 1-511 1-503 1-486 2-246 2-233 2-240 2-241 2-260 2-283 2-285 2-288 2-295 2-296 2-297 2-300 2-544 2-554 2-570 3-244 3-284 3-239 3-290 3-291 3-292 3-294 3-243 3-238 4-256 4-257 4-259 4-287 4-251 4-250 4-255 5-299 5-274 6-280 6-270 6-248 6-262 6-261 6-275 6-277 6-281 7-267 7-254 7-273 7-258 7-266 7-279 7-286 8-253 8-269 8-289 8-272 8-282 9-278 9-263 10-271 10-268 11-276 12-264 19-265</p>

COLUMN NAME	DATATYPE	COMMENTS
 id	serial	min : 1 max : 2000000
 quantity	real	min : 1 max : 60 avg : 30.5135215 dolný kvartil : 16 medián : 30 horný kvartil : 46
 stuff_id	integer	min : 1 max : 100000 avg : 49976.5378755 dolný kvartil : 24967 medián : 49955 horný kvartil : 74990.25
 service_id	integer	min : 1 max : 2000000 avg : 1000000.5 dolný kvartil : 500000.75 medián : 1000000.5 horný kvartil : 1500000.25



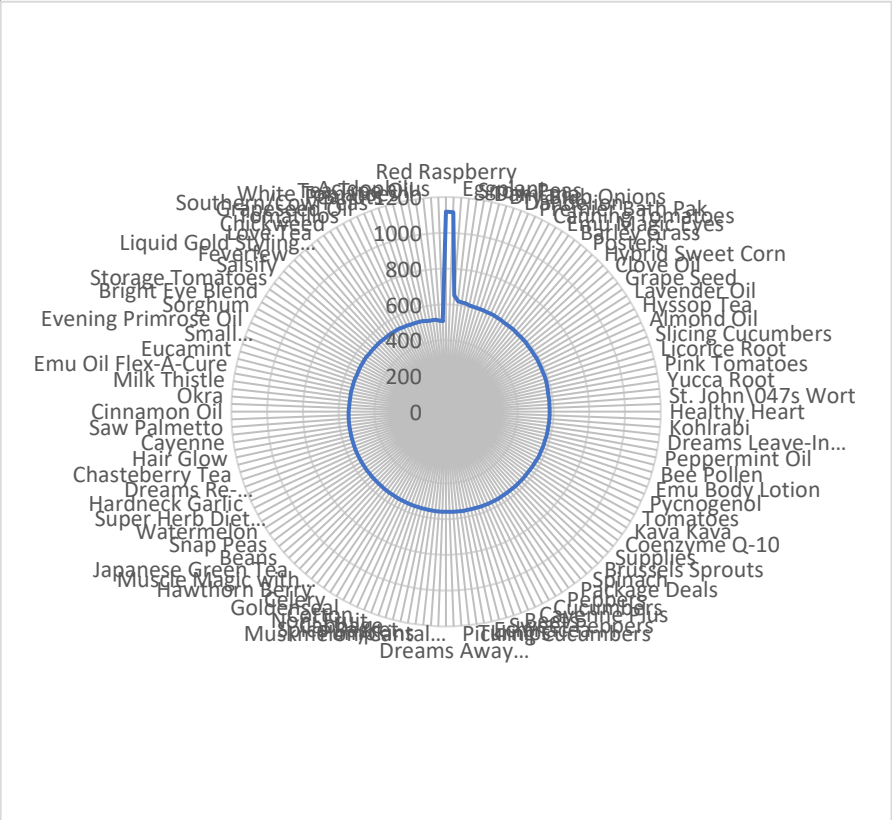

service


count: 2 000 000

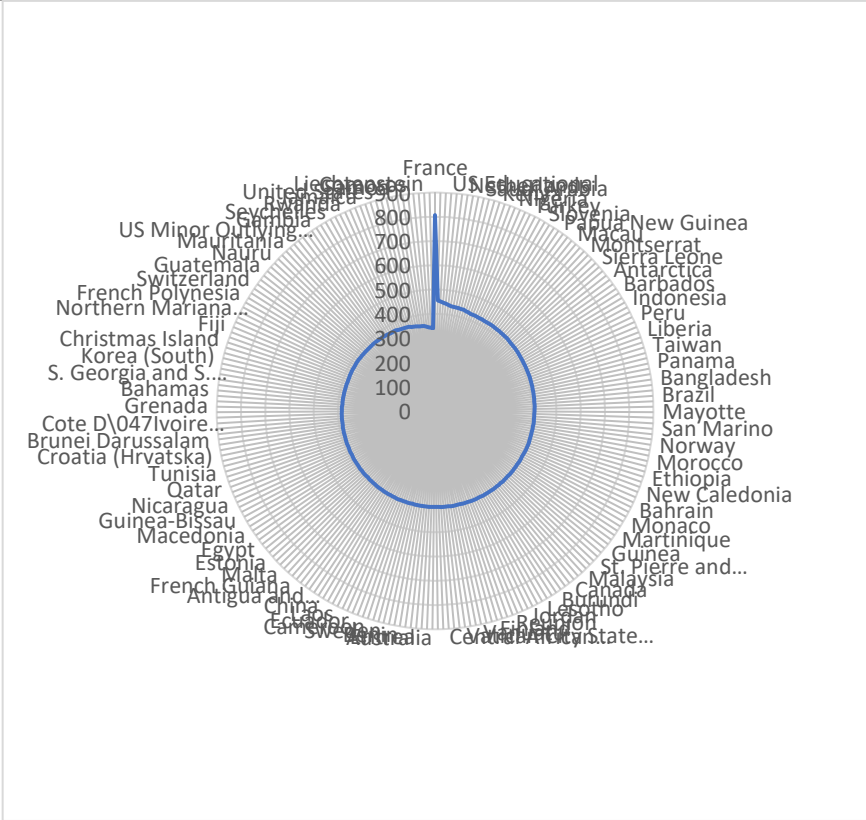

COLUMN NAME	DATATYPE	COMMENTS
 id	serial	min : 1 max : 2000000

COLUMN NAME	DATATYPE	COMMENTS
 executed_date	date	min : 1990 max : 2017 avg : 2003.3085575 dolný kvartil : 1996 medián : 2003 horný kvartil : 2010
 cost	real	min : 1 max : 300 avg : 105.2526035 dolný kvartil : 47 medián : 79 horný kvartil : 151
 invoice_id	integer	min : 1 max : 1000000 avg : 500000.5 dolný kvartil : 250000.75 medián : 500000.5 horný kvartil : 750000.25
 type_id	integer	min : 1 max : 80 avg : 18.868081 dolný kvartil : 6 medián : 16 horný kvartil : 24

count: 100 000

COLUMN NAME	DATATYPE	COMMENTS
 id	serial	min : 1 max : 100000
 name	varchar(50)	<p>the number of repeating values - how many times :</p> <p>1-607 1-605 1-507 1-601 1-519 1-592 1-584 1-524 1-583 1-582 1-560 1-510 1-536 1-573 1-1119 1-1116 1-651 1-622 1-618 1-529 1-544 1-530 1-616 1-608 1-532 2-604 2-525 2-526 2-533 2-534 2-535 2-537 2-543 2-545 2-548 2-553 2-556 2-561 2-562 2-564 2-567 2-569 2-570 2-572 2-578 2-586 2-590 2-591 2-598 2-602 3-577 3-587 3-588 3-554 3-538 3-518 3-565 3-566 3-549 3-563 3-575 3-576 3-580 3-559 3-555 3-585 3-542 4-541 4-552 4-568 4-546 4-547 4-558 4-594 4-551 5-579 6-550 7-540</p> 
 brand	varchar(50)	<p>the number of repeating values - how many times :</p> <p>1-692 1-691 1-687 1-685 1-683 1-721 1-681 1-598 1-680 1-609 1-628 1-593 1-679 1-677 1-676 1-673 1-661 1-671 1-630 1-631 1-670 1-709 1-633 1-705 1-701 1-722 1-637 1-639 1-698 1-614 1-694 2-629 2-592 2-605 2-606 2-612 2-615 2-616 2-621 2-623 2-626 2-627 2-590 2-636 2-641 2-648 2-651 2-654 2-655 2-656 2-664 2-666 2-669 2-674 2-675 2-689 2-693 2-697 3-635 3-659 3-640 3-667 3-672 3-660 3-652 3-649 3-647 3-646 3-645 3-632 3-644 3-643 3-624 3-642 3-622 4-658 4-662 5-657 5-634</p>

COLUMN NAME	DATATYPE	COMMENTS
 supplier_ name	varchar(50)	<p>the number of repeating values - how many times :</p> <p>1-447 1-444 1-440 1-436 1-434 1-430 1-429 1-420 1-358 1-359 1-342 1-345 1-365 1-366 1-807 1-370 1-371 1-349 1-352 1-374 1-455 1-343 1-355 1-357 1-452 2-363 2-354 2-364 2-372 2-373 2-376 2-377 2-378 2-406 2-407 2-410 2-415 2-417 2-419 2-424 2-425 2-426 2-433 2-435 3-381 3-412 3-384 3-422 3-382 3-418 3-395 3-416 4-399 4-405 4-375 4-413 4-369 4-383 4-421 4-388 4-391 4-392 4-397 5-404 5-403 5-387 5-386 5-400 5-389 5-409 5-411 5-414 5-398 5-402 5-423 5-401 5-379 6-390 6-385 6-393 7-380 8-394 8-396 9-408</p>

COLUMN NAME	DATATYPE	COMMENTS
		
 available_quantity	real	<p>min : 30</p> <p>max : 600</p> <p>avg : 314.35076</p> <p>dolný kvartil : 171</p> <p>medián : 314</p> <p>horný kvartil : 457</p>

type

count: 23

COLUMN NAME	DATATYPE	COMMENTS
 id	serial	<div>min : 1</div> <div>max : 80</div> <div>avg : 18.868081</div> <div>dolný kvartil : 6</div> <div>medián : 16</div>

COLUMN NAME	DATATYPE	COMMENTS
		horný kvartil : 24
label	varchar(50)	<p>the number of repeating values - how many times : 23-1</p>