

# Projekt BigData

Michał Degowski s16584

System analizujący dane StackOverflow pobrane jako zbiór danych Google BigQuery. Ma na celu poznanie trendów technologicznych od 2012 do 2021 roku i przewidzenie przyszłych trendów technologicznych.

## 1. Oprogramowanie i biblioteki, które wykorzystuję na dany moment :

- kaggle notebook – edytor tekstu;
- Python - język programowania;
- Numpy – biblioteka dodająca obsługę dużych, wielowymiarowych tablic i macierzy, a także duży zbiór funkcji matematycznych wysokiego poziomu do obsługi tych tablic;
- Pandas – biblioteka do manipulacji i analizy danych;
- Matplotlib - biblioteka do tworzenia wykresów;
- Pandas Profiling – moduł za pomocą którego możemy szybko przeprowadzić eksploracyjną analizę danych;
- sklearn - biblioteka do uczenia maszynowego
- statsmodels - pozwala użytkownikom eksplorować dane, szacować modele statystyczne i przeprowadzać testy statystyczne
- wordcloud - generator chmury słów

## 2. Wykorzystywane dane:

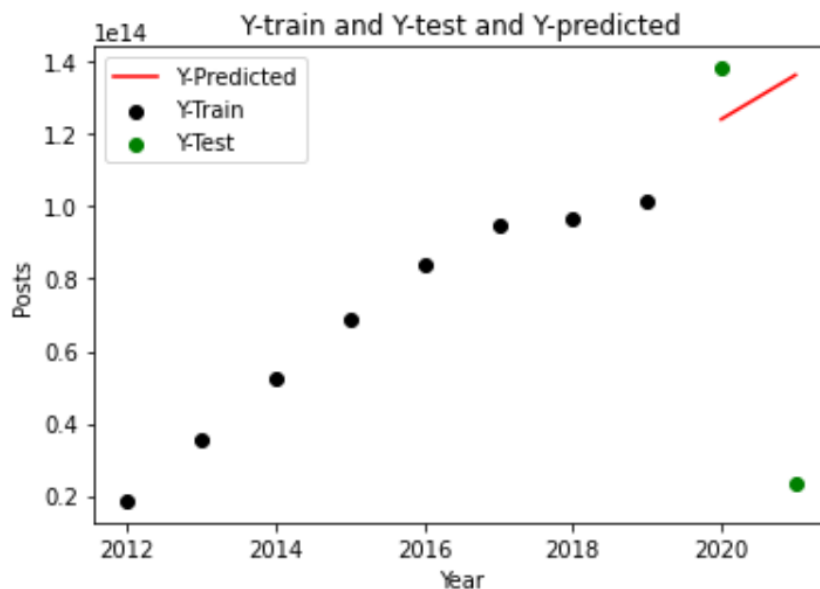
<https://console.cloud.google.com/marketplace/product/stack-exchange/stack-overflow?project=ony-x-link-317010&folder=&organizationId=>

## 3. Ogólny zarys projektu:

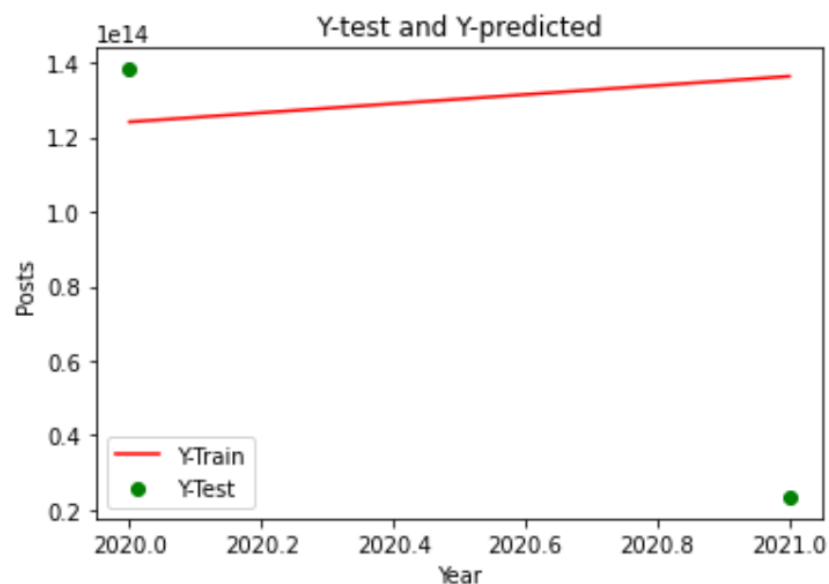
- 1) Jaki jest trend technologiczny w latach 2012 - 2021?
- 2) Jaki jest trend w różnych dziedzinach każdej technologii?
- 3) Jakiego będą nadchodzące trendy technologiczne?
- 4) Jaki będzie nadchodzący trend w różnych dziedzinach każdej technologii?

## 4. Wykresy i wnioski:

Wykres wizualizujący train data, test data oraz predicted dla całego modelu



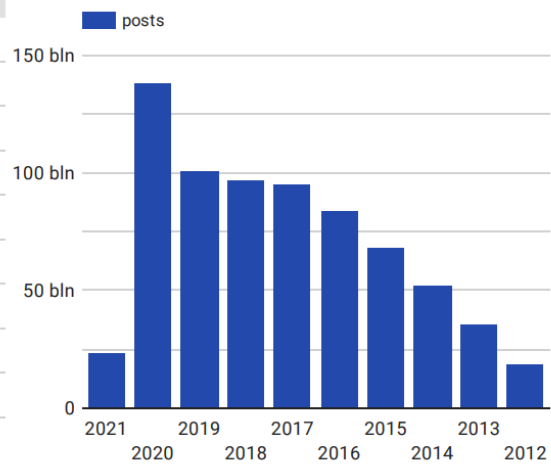
Wykres wizualizujący test values i predicted values w celu sprawdzenia dokładności modelu



### Liczba postów na przestrzeni lat 2012-2021

	year	posts
1.	2021	23 465 045 504 735
2.	2020	138 196 879 651 125
3.	2019	101 305 194 030 008
4.	2018	96 903 239 286 068
5.	2017	94 916 561 257 721
6.	2016	83 815 179 254 633
7.	2015	68 658 879 146 492
8.	2014	52 223 931 188 088
9.	2013	35 741 489 445 223
10.	2012	18 625 782 310 047

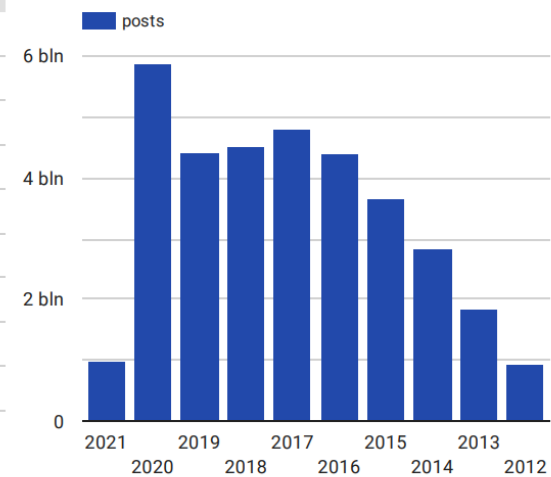
1 - 10 / 10 < >



### Liczba postów WebDev na przestrzeni lat 2012-2021

	year	posts
1.	2021	4 606 682 737 909
2.	2020	26 961 572 282 026
3.	2019	20 502 880 294 776
4.	2018	21 017 472 215 283
5.	2017	22 878 391 507 007
6.	2016	21 182 335 348 268
7.	2015	17 639 654 830 048
8.	2014	13 636 184 053 085
9.	2013	9 070 061 991 148
10.	2012	4 454 577 591 437

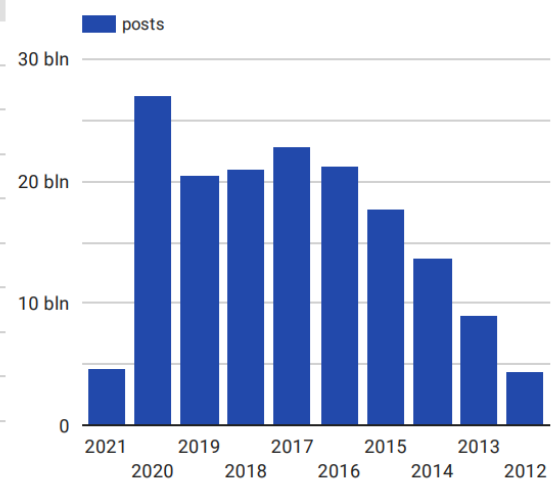
1 - 10 / 10 < >



### Liczba postów DataBase na przestrzeni lat 2012-2021

	year	posts
1.	2021	971 433 713 240
2.	2020	5 880 532 229 527
3.	2019	4 423 724 457 674
4.	2018	4 517 019 984 154
5.	2017	4 802 811 677 522
6.	2016	4 384 155 123 277
7.	2015	3 657 623 642 722
8.	2014	2 825 811 484 215
9.	2013	1 844 090 100 234
10.	2012	924 589 942 903

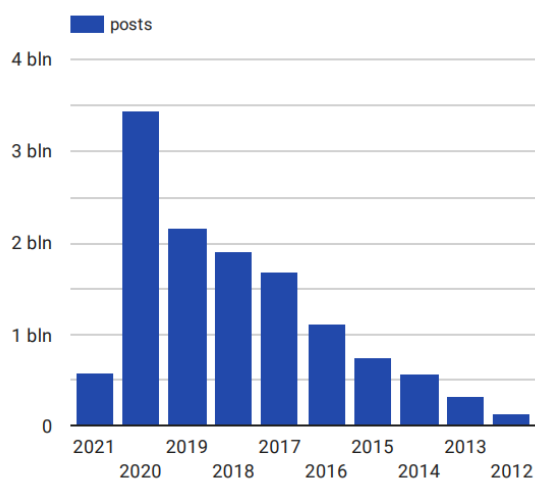
1 - 10 / 10 < >



### Liczba postów DataScience na przestrzeni lat 2012-2021

	year	posts
1.	2021	578 635 330 720
2.	2020	3 432 446 731 872
3.	2019	2 162 040 512 121
4.	2018	1 907 737 135 653
5.	2017	1 690 770 598 001
6.	2016	1 121 208 548 783
7.	2015	751 868 448 936
8.	2014	562 933 239 869
9.	2013	325 613 762 891
10.	2012	127 026 121 714

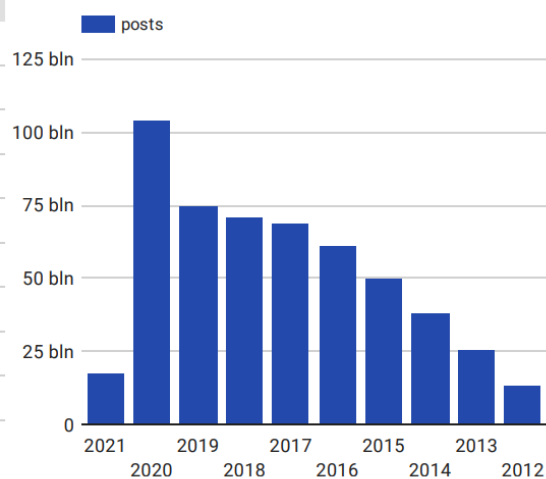
1 - 10 / 10 < >



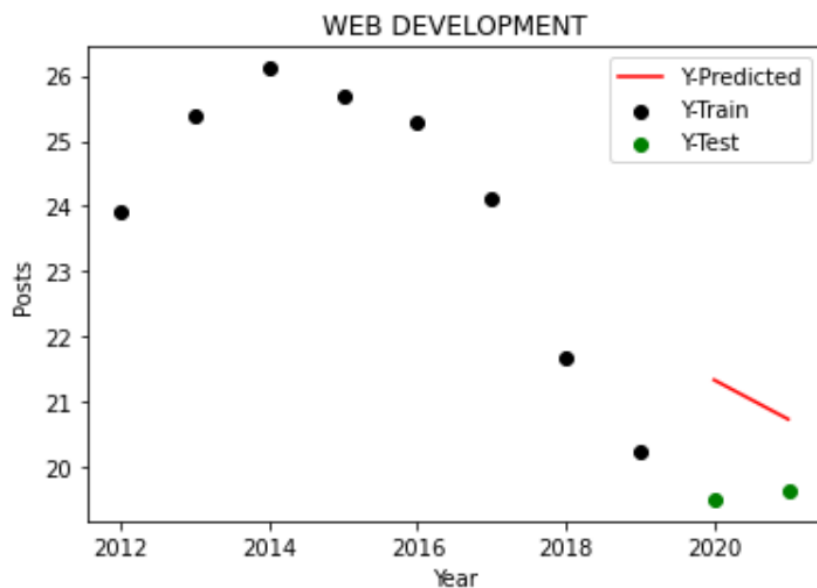
### Liczba postów ProgLang na przestrzeni lat 2012-2021

	year	posts
1.	2021	17 639 301 298 979
2.	2020	103 819 520 968 773
3.	2019	74 679 020 997 764
4.	2018	70 916 925 765 751
5.	2017	69 298 394 234 892
6.	2016	60 954 101 616 486
7.	2015	50 127 960 391 458
8.	2014	38 041 145 561 813
9.	2013	25 854 595 415 324
10.	2012	13 388 811 321 750

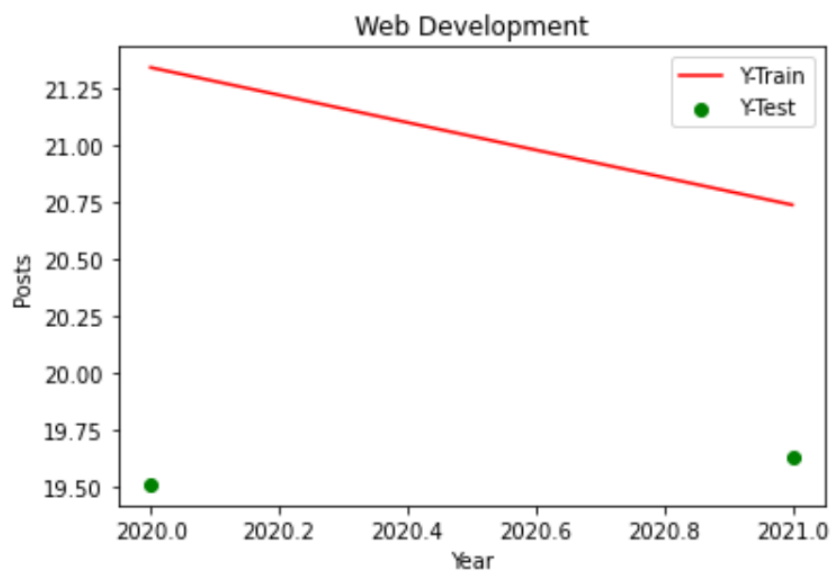
1 - 10 / 10 < >



Wykres wizualizujący train data, test data oraz predicted dla Web Development



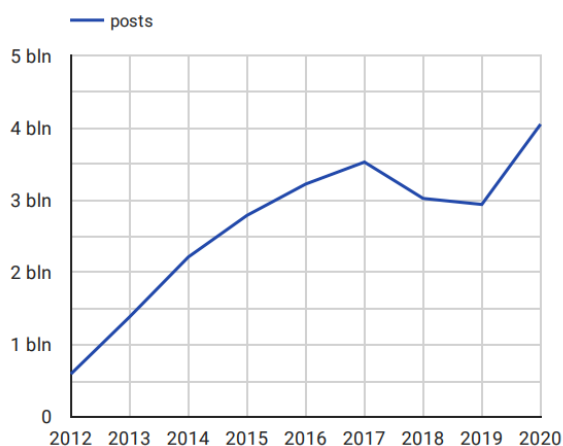
Wykres wizualizujący test values i predicted values w celu sprawdzenia dokładności modelu



### Liczba postów z tagiem 'CSS' na przestrzeni lat 2012-2021

	year	posts
1.	2021	710 002 913 296
2.	2020	4 051 368 936 620
3.	2019	2 941 775 622 614
4.	2018	3 024 949 409 367
5.	2017	3 528 066 232 427
6.	2016	3 224 047 630 076
7.	2015	2 790 578 136 875
8.	2014	2 214 243 378 059
9.	2013	1 386 294 153 637
10.	2012	595 880 917 231

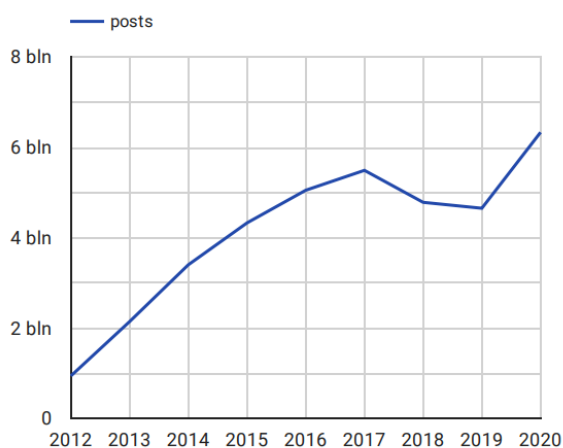
1 - 10 / 10 < >



### Liczba postów z tagiem 'HTML' na przestrzeni lat 2012-2021

	year	posts
1.	2021	1 059 221 856 956
2.	2020	6 331 235 645 455
3.	2019	4 650 130 142 677
4.	2018	4 781 277 930 111
5.	2017	5 489 165 371 019
6.	2016	5 048 711 046 314
7.	2015	4 325 859 157 466
8.	2014	3 398 987 090 158
9.	2013	2 141 306 846 597
10.	2012	938 881 319 676

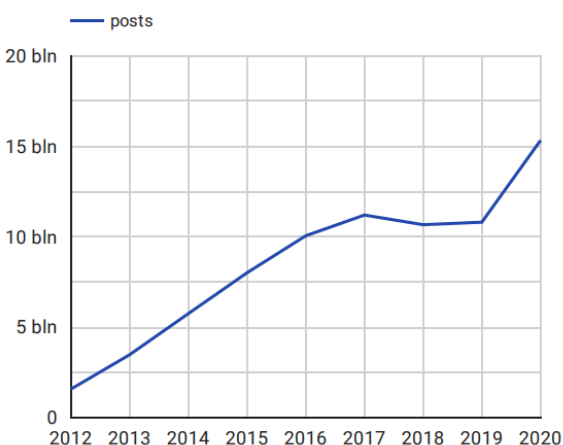
1 - 10 / 10 < >



### Liczba postów z tagiem 'JavaScript' na przestrzeni lat 2012-2021

	year	posts
1.	2021	2 632 442 511 985
2.	2020	15 341 995 594 929
3.	2019	10 812 168 417 203
4.	2018	10 673 607 991 142
5.	2017	11 206 909 763 043
6.	2016	10 060 816 534 669
7.	2015	8 012 401 215 573
8.	2014	5 752 342 148 464
9.	2013	3 477 506 590 732
10.	2012	1 573 430 871 489

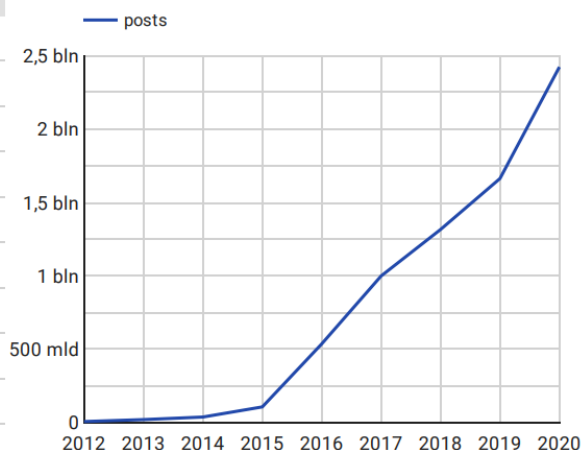
1 - 10 / 10 < >



### Liczba postów z tagiem 'TypeScript' na przestrzeni lat 2012-2021

	year	posts
1.	2021	415 405 620 074
2.	2020	2 425 360 259 551
3.	2019	1 664 328 011 830
4.	2018	1 316 769 109 384
5.	2017	999 805 022 277
6.	2016	536 446 376 232
7.	2015	106 444 156 351
8.	2014	37 370 833 799
9.	2013	19 964 855 747
10.	2012	6 590 856 918

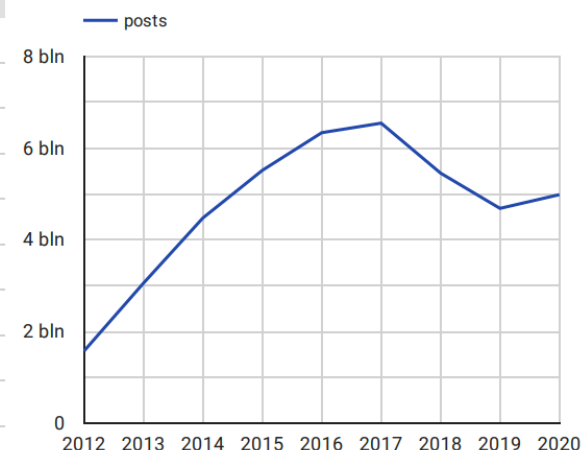
1 - 10 / 10 < >



### Liczba postów z tagiem 'PHP' na przestrzeni lat 2012-2021

	year	posts
1.	2021	853 791 185 283
2.	2020	4 981 612 768 476
3.	2019	4 681 629 691 677
4.	2018	5 451 982 697 230
5.	2017	6 542 127 383 272
6.	2016	6 333 918 241 521
7.	2015	5 511 726 842 367
8.	2014	4 472 618 948 455
9.	2013	3 049 522 271 889
10.	2012	1 568 939 578 850

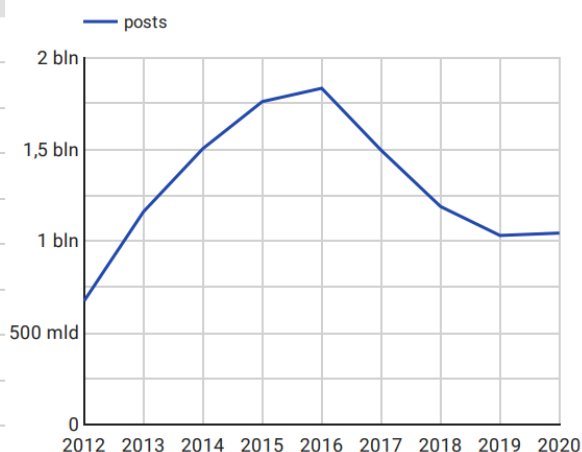
1 - 10 / 10 < >



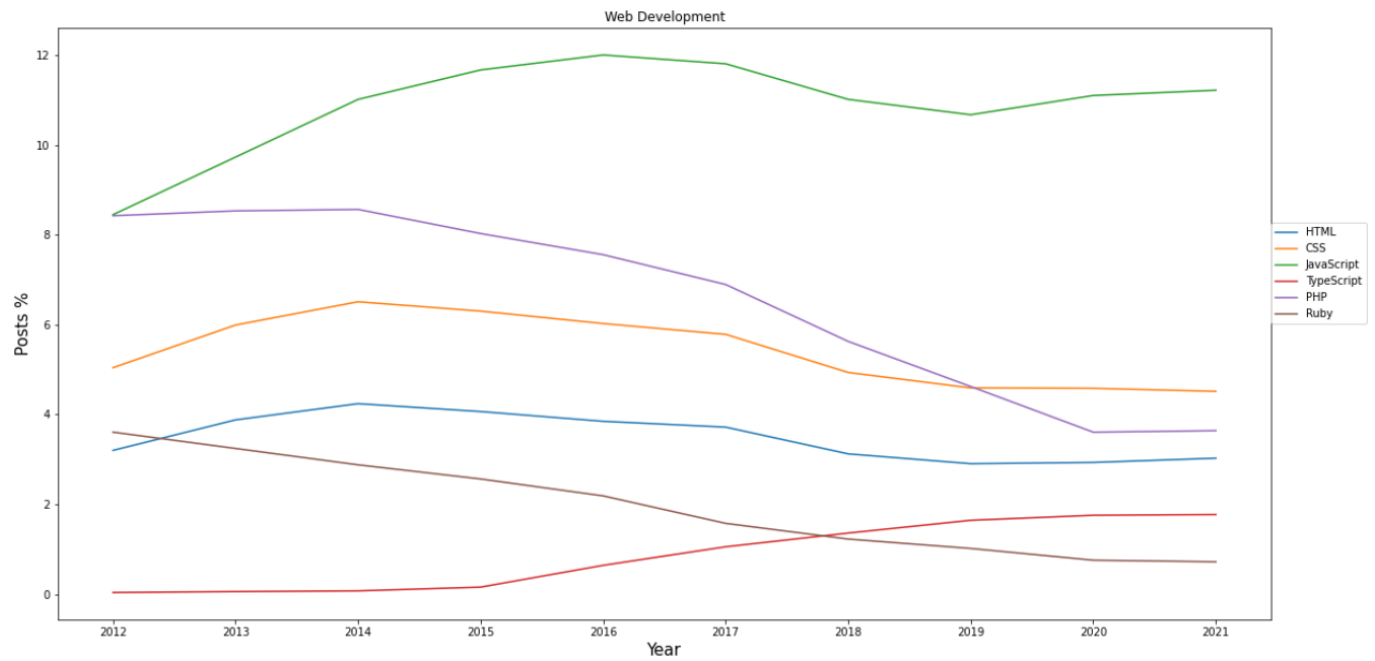
### Liczba postów z tagiem 'Ruby' na przestrzeni lat 2012-2021

	year	posts
1.	2021	168 831 393 681
2.	2020	1 042 240 074 887
3.	2019	1 029 226 361 954
4.	2018	1 187 068 376 798
5.	2017	1 494 292 117 653
6.	2016	1 832 506 574 592
7.	2015	1 759 812 174 652
8.	2014	1 503 144 174 377
9.	2013	1 158 648 852 876
10.	2012	671 594 421 262

1 - 10 / 10 < >

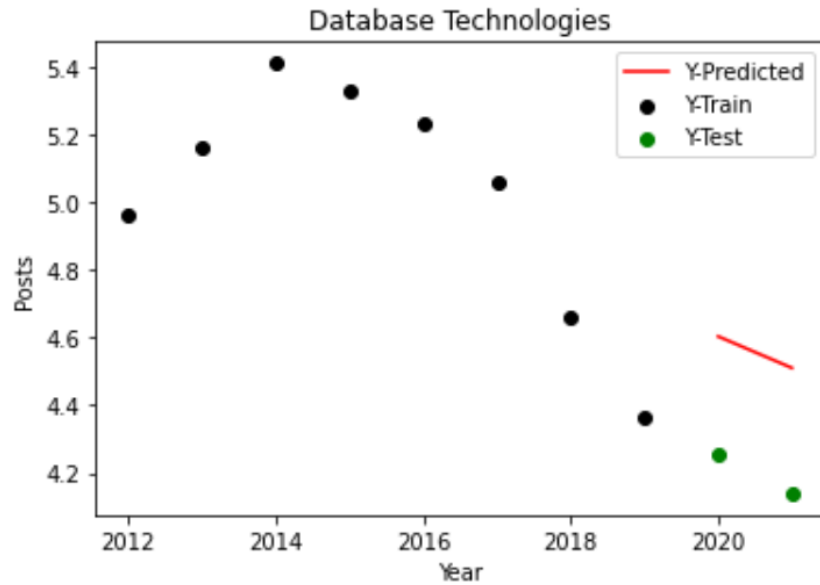


Wykres porównujący popularności różnych kategorii w ramach Web Development

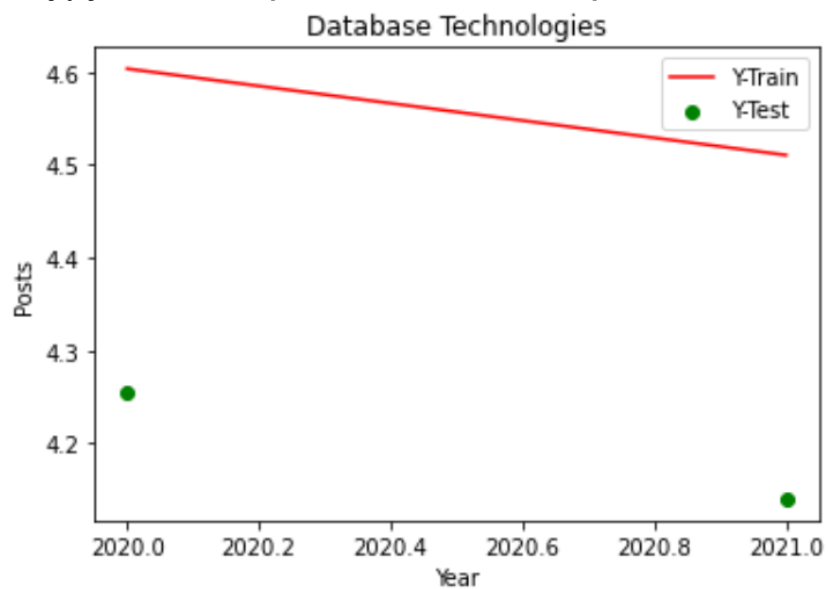




Wykres wizualizujący train data, test data oraz predicted dla Database



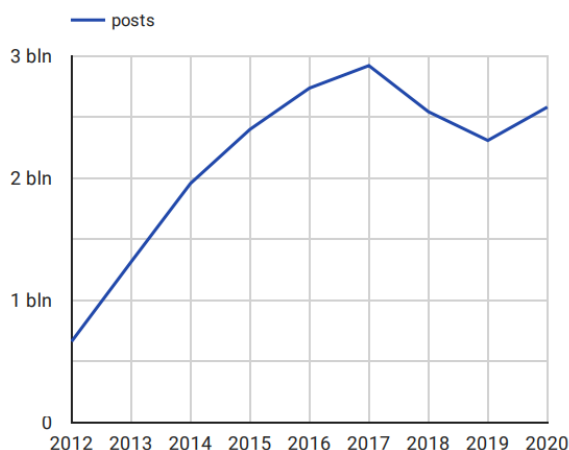
Wykres wizualizujący test values i predicted values w celu sprawdzenia dokładności modelu



### Liczba postów z tagiem 'MySQL' na przestrzeni lat 2012-2021

	year	posts
1.	2021	411 362 181 707
2.	2020	2 581 476 124 264
3.	2019	2 308 565 320 007
4.	2018	2 543 904 123 249
5.	2017	2 921 248 597 580
6.	2016	2 737 012 386 274
7.	2015	2 399 109 752 133
8.	2014	1 956 567 953 432
9.	2013	1 313 208 701 670
10.	2012	662 415 724 137

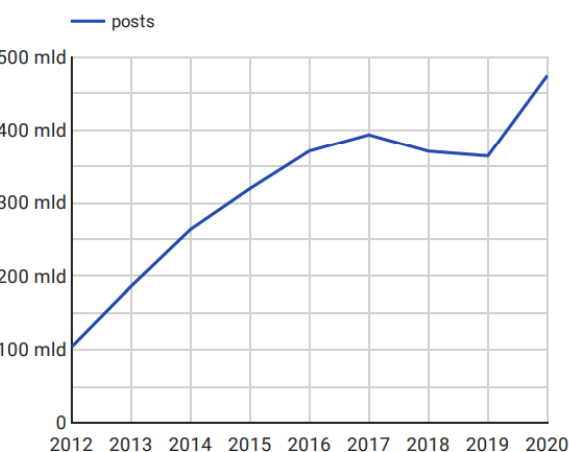
1 - 10 / 10 < >



### Liczba postów z tagiem 'SQLite' na przestrzeni lat 2012-2021

	year	posts
1.	2021	80 286 743 705
2.	2020	474 819 868 429
3.	2019	364 136 237 902
4.	2018	370 582 646 087
5.	2017	393 597 202 885
6.	2016	370 940 915 465
7.	2015	319 356 040 471
8.	2014	263 565 956 645
9.	2013	185 950 301 578
10.	2012	103 392 607 127

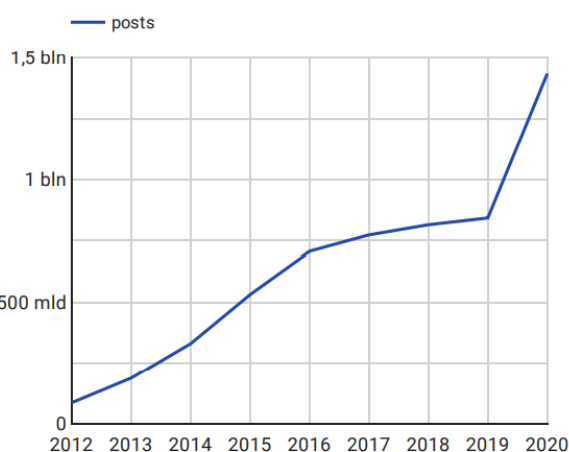
1 - 10 / 10 < >



### Liczba postów z tagiem 'MongoDB' na przestrzeni lat 2012-2021

	year	posts
1.	2021	236 638 737 299
2.	2020	1 434 611 381 521
3.	2019	841 983 170 233
4.	2018	814 113 745 640
5.	2017	772 867 646 005
6.	2016	706 122 291 562
7.	2015	531 559 914 648
8.	2014	330 653 304 752
9.	2013	187 059 145 325
10.	2012	85 429 968 843

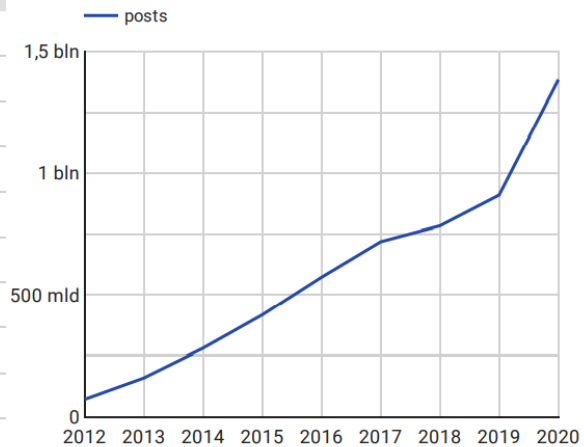
1 - 10 / 10 < >



### Liczba postów z tagiem 'PostgreSQL' na przestrzeni lat 2012-2021

	year	posts
1.	2021	240 175 911 250
2.	2020	1 383 910 743 919
3.	2019	909 146 901 018
4.	2018	783 282 371 613
5.	2017	718 836 306 691
6.	2016	572 952 772 791
7.	2015	416 912 266 344
8.	2014	280 090 213 093
9.	2013	159 939 310 829
10.	2012	72 362 835 773

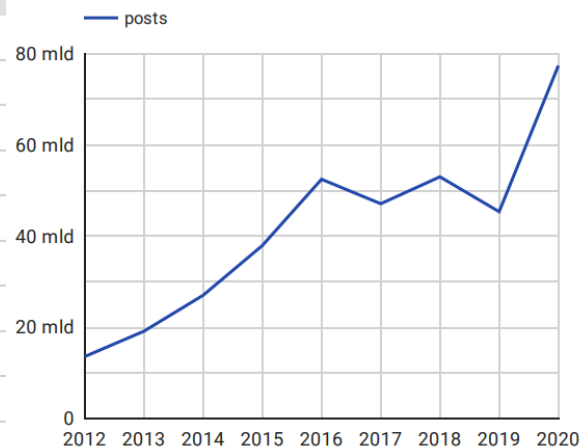
1 - 10 / 10 < >



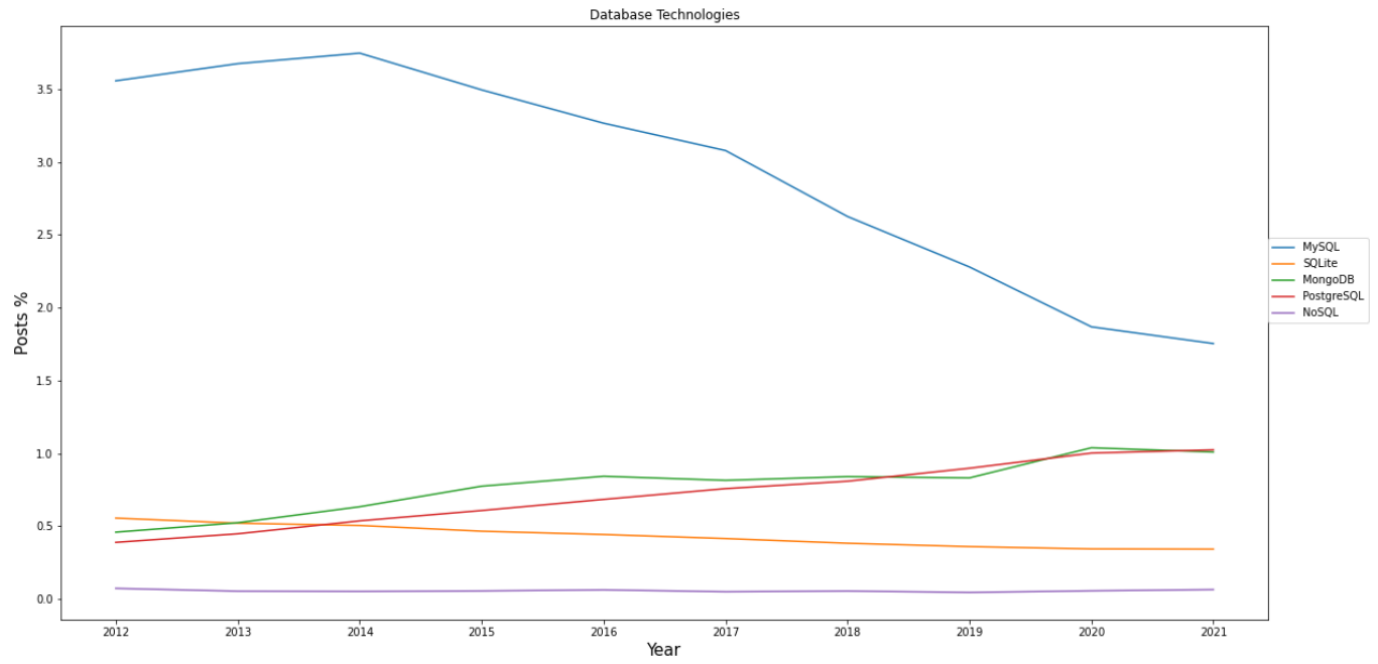
### Liczba postów z tagiem 'NoSQL' na przestrzeni lat 2012-2021

	year	posts
1.	2021	15 114 290 678
2.	2020	77 387 574 933
3.	2019	45 324 952 857
4.	2018	52 981 576 481
5.	2017	47 076 891 280
6.	2016	52 444 787 830
7.	2015	37 897 299 693
8.	2014	26 984 414 052
9.	2013	19 095 044 790
10.	2012	13 531 435 622

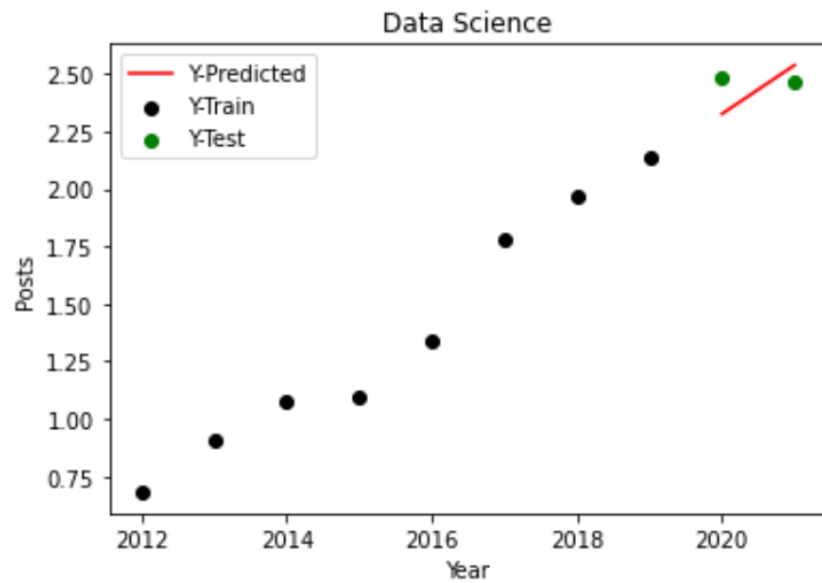
1 - 10 / 10 < >



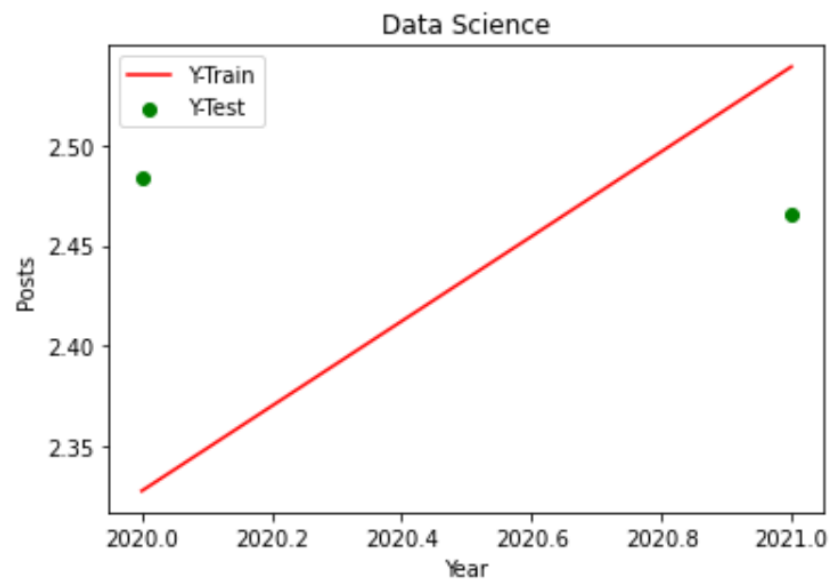
Wykres porównujący popularności różnych kategorii w ramach Database



Wykres wizualizujący train data, test data oraz predicted dla Data Science

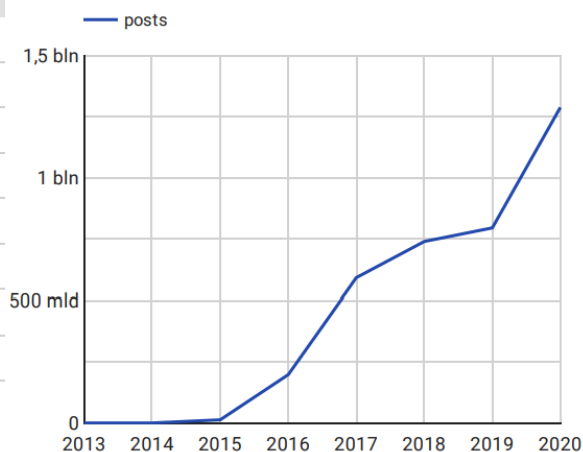


Wykres wizualizujący test values i predicted values w celu sprawdzenia dokładności modelu



### Liczba postów z tagiem 'TensorFlow' na przestrzeni lat 2013-2021

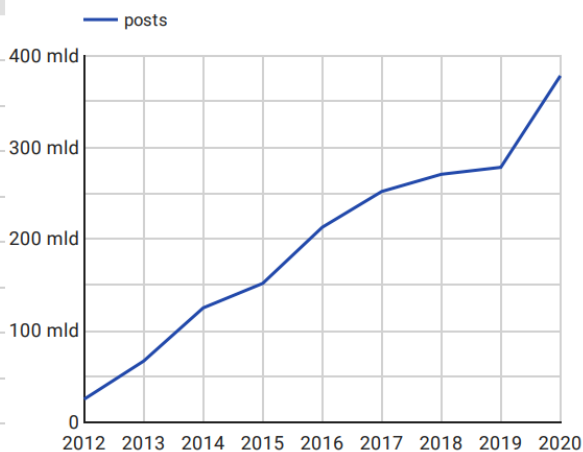
	year	posts
1.	2021	206 654 214 605
2.	2020	1 288 065 046 152
3.	2019	794 697 945 172
4.	2018	738 964 743 584
5.	2017	590 946 007 707
6.	2016	197 880 414 440
7.	2015	12 991 888 815
8.	2013	38 055 340



1 - 8 / 8 < >

### Liczba postów z tagiem 'SAS' na przestrzeni lat 2012-2021

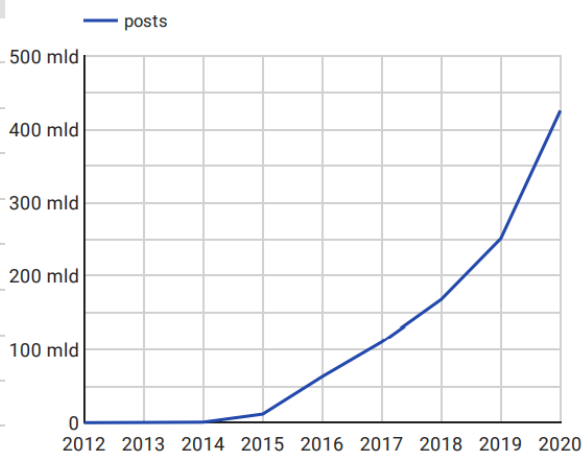
	year	posts
1.	2021	60 955 242 821
2.	2020	377 920 981 923
3.	2019	277 969 414 495
4.	2018	270 361 950 258
5.	2017	251 595 976 610
6.	2016	212 498 822 964
7.	2015	151 231 120 704
8.	2014	124 479 650 549
9.	2013	66 436 637 563
10.	2012	24 751 664 819



1 - 10 / 10 < >

### Liczba postów z tagiem 'Jupyter' na przestrzeni lat 2012-2021

	year	posts
1.	2021	78 049 892 373
2.	2020	426 139 623 914
3.	2019	251 541 778 852
4.	2018	168 524 897 532
5.	2017	110 542 050 451
6.	2016	63 097 187 225
7.	2015	11 879 933 067
8.	2014	832 552 534
9.	2013	430 059 052
10.	2012	124 328 295

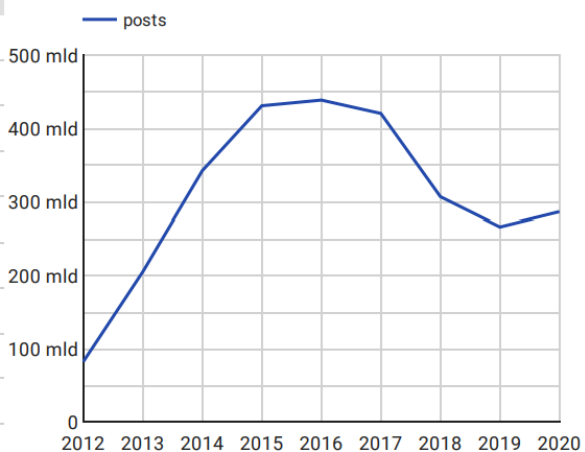


1 - 10 / 10 < >

### Liczba postów z tagiem 'MATLAB' na przestrzeni lat 2012-2021

	year	posts
1.	2021	54 309 823 912
2.	2020	287 093 891 537
3.	2019	267 158 365 255
4.	2018	307 452 592 662
5.	2017	420 878 906 681
6.	2016	439 086 853 322
7.	2015	431 363 787 505
8.	2014	342 847 893 693
9.	2013	206 229 989 423
10.	2012	83 640 927 944

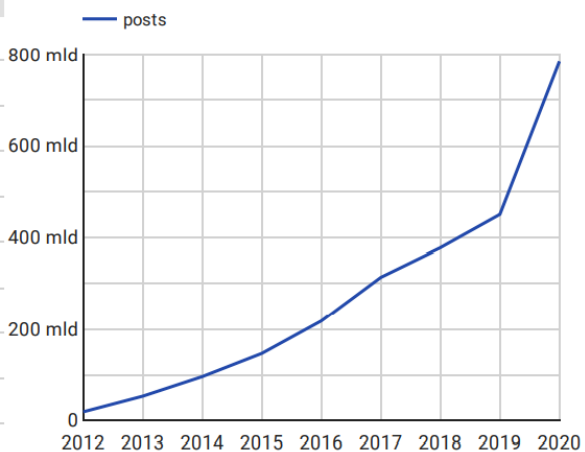
1 - 10 / 10 < >



### Liczba postów z tagiem 'Matplotlib' na przestrzeni lat 2012-2021

	year	posts
1.	2021	127 021 188 353
2.	2020	783 756 798 822
3.	2019	449 801 353 192
4.	2018	377 406 487 542
5.	2017	314 111 163 202
6.	2016	217 684 630 843
7.	2015	146 372 353 174
8.	2014	95 617 426 436
9.	2013	52 892 371 861
10.	2012	18 676 450 426

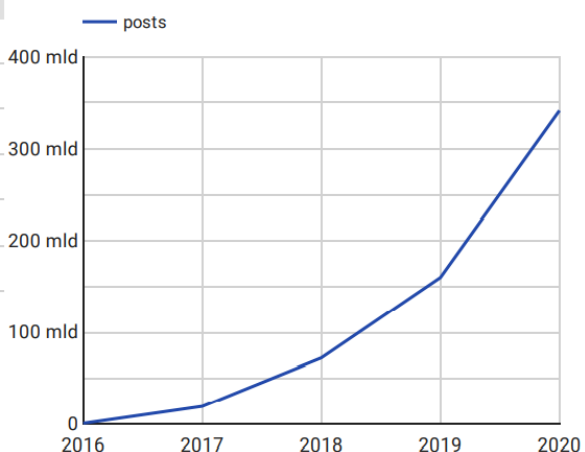
1 - 10 / 10 < >



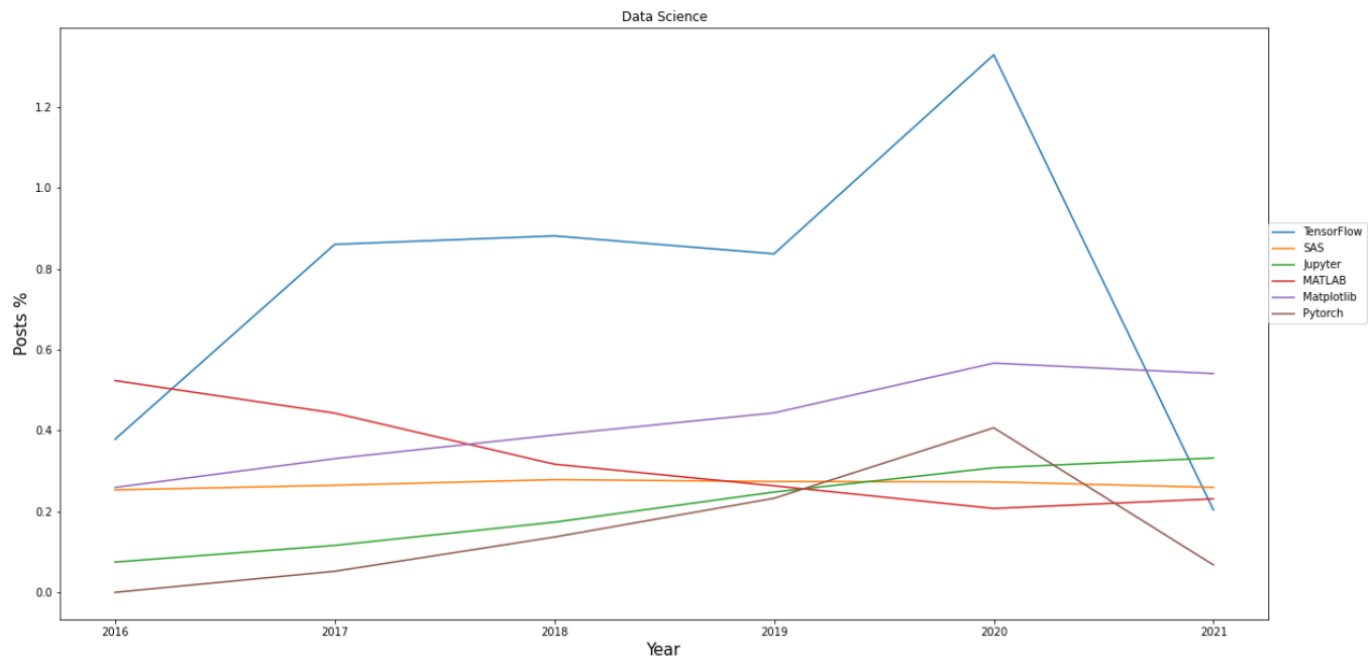
### Liczba postów z tagiem 'Pytorch' na przestrzeni lat 2016-2021

	year	posts
1.	2021	64 709 958 248
2.	2020	341 275 654 238
3.	2019	159 954 637 944
4.	2018	71 621 279 696
5.	2017	18 772 019 353
6.	2016	73 294 118

1 - 6 / 6 < >

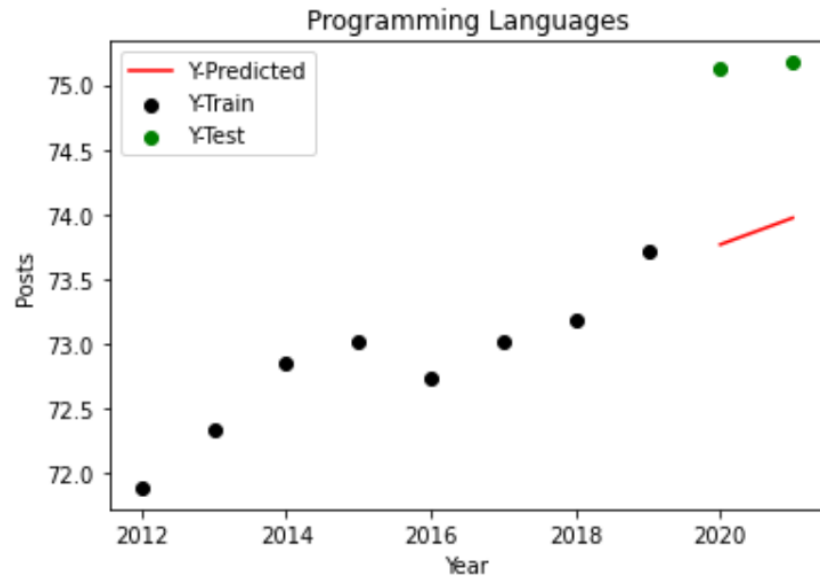


Wykres porównujący popularności różnych kategorii w ramach Data Science

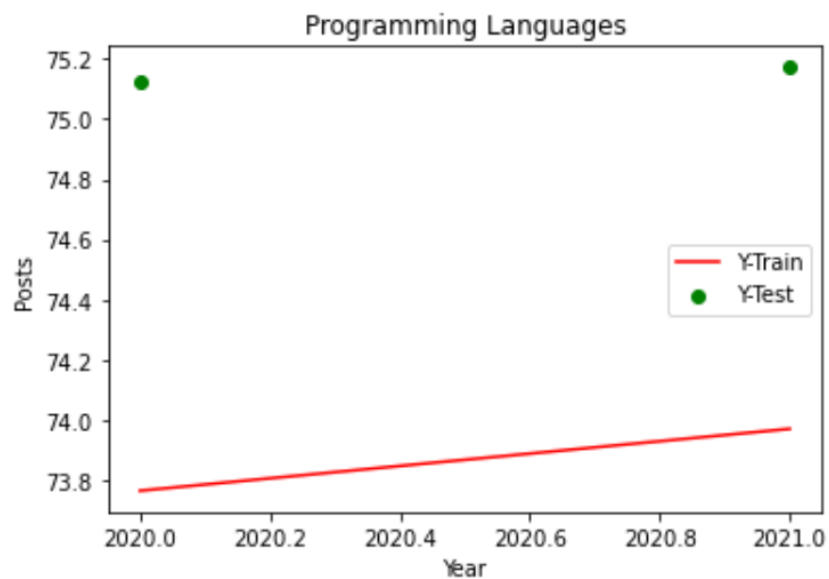




Wykres wizualizujący train data, test data oraz predicted dla Data Science



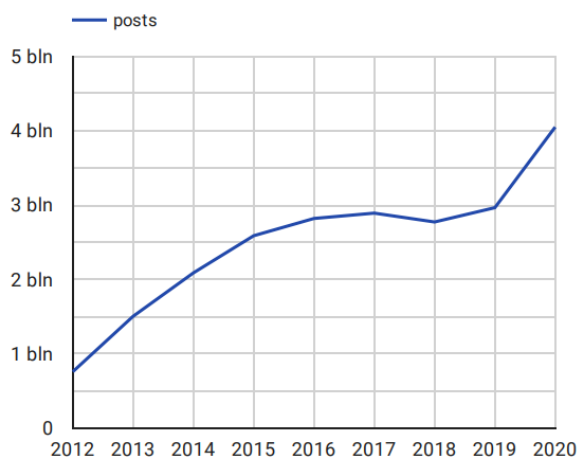
Wykres wizualizujący test values i predicted values w celu sprawdzenia dokładności modelu



### Liczba postów z tagiem 'C++' na przestrzeni lat 2012-2021

	year	posts
1.	2021	654 543 881 943
2.	2020	4 049 914 228 775
3.	2019	2 964 723 984 101
4.	2018	2 772 212 223 326
5.	2017	2 891 345 979 766
6.	2016	2 819 435 190 637
7.	2015	2 586 874 453 066
8.	2014	2 086 623 929 663
9.	2013	1 502 075 714 758
10.	2012	755 719 005 301

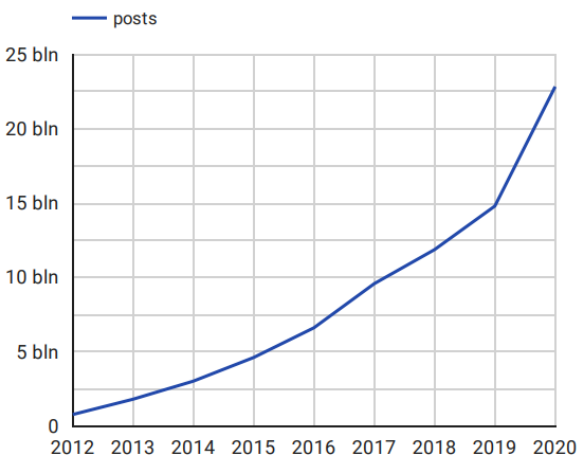
1 - 10 / 10 < >



### Liczba postów z tagiem 'Python' na przestrzeni lat 2012-2021

	year	posts
1.	2021	3 979 424 741 946
2.	2020	22 834 897 847 516
3.	2019	14 795 711 538 366
4.	2018	11 865 292 180 963
5.	2017	9 572 904 481 949
6.	2016	6 604 463 503 213
7.	2015	4 603 747 943 826
8.	2014	3 012 329 715 085
9.	2013	1 796 019 684 557
10.	2012	766 382 707 792

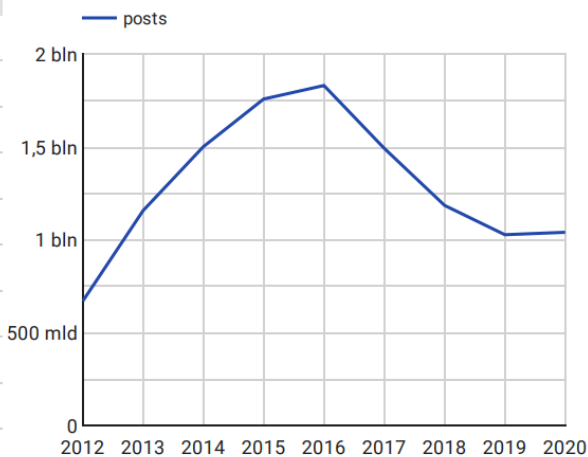
1 - 10 / 10 < >



### Liczba postów z tagiem 'Ruby' na przestrzeni lat 2012-2021

	year	posts
1.	2021	168 831 393 681
2.	2020	1 042 240 074 887
3.	2019	1 029 226 361 954
4.	2018	1 187 068 376 798
5.	2017	1 494 292 117 653
6.	2016	1 832 506 574 592
7.	2015	1 759 812 174 652
8.	2014	1 503 144 174 377
9.	2013	1 158 648 852 876
10.	2012	671 594 421 262

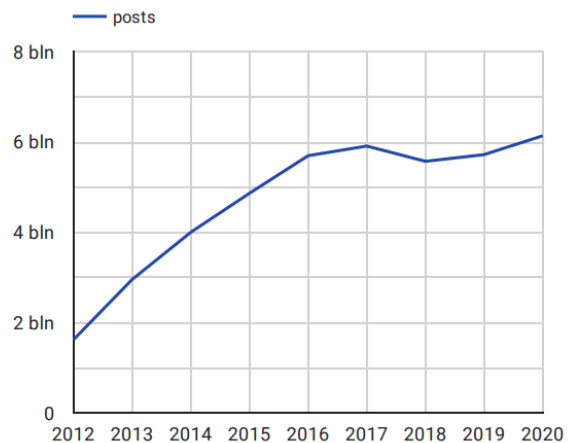
1 - 10 / 10 < >



### Liczba postów z tagiem 'C#' na przestrzeni lat 2012-2021

	year	posts
1.	2021	983 669 814 632
2.	2020	6 141 983 442 561
3.	2019	5 724 245 360 344
4.	2018	5 576 256 313 788
5.	2017	5 914 090 941 261
6.	2016	5 700 429 878 541
7.	2015	4 868 195 413 380
8.	2014	4 004 631 210 007
9.	2013	2 958 689 164 060
10.	2012	1 624 587 422 374

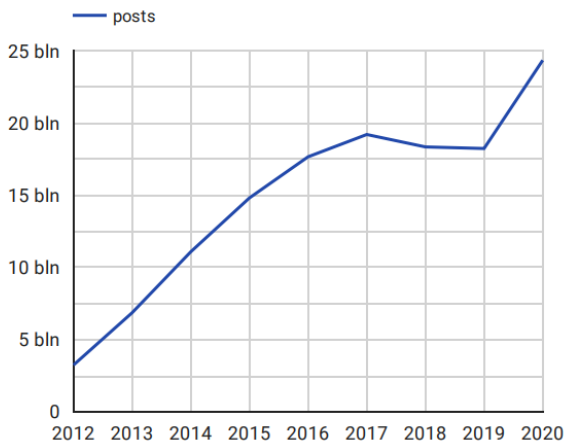
1 - 10 / 10 < >



### Liczba postów z tagiem 'Java' na przestrzeni lat 2012-2021

	year	posts
1.	2021	3 996 824 770 638
2.	2020	24 350 443 501 546
3.	2019	18 240 210 652 310
4.	2018	18 350 847 562 967
5.	2017	19 209 752 980 895
6.	2016	17 667 284 143 060
7.	2015	14 820 975 575 066
8.	2014	11 092 275 098 589
9.	2013	6 873 799 512 073
10.	2012	3 250 981 167 435

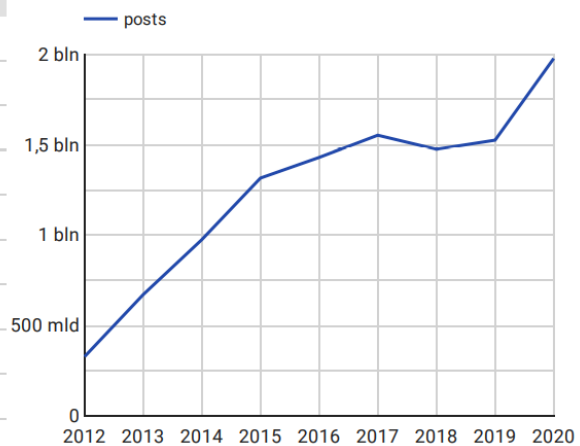
1 - 10 / 10 < >



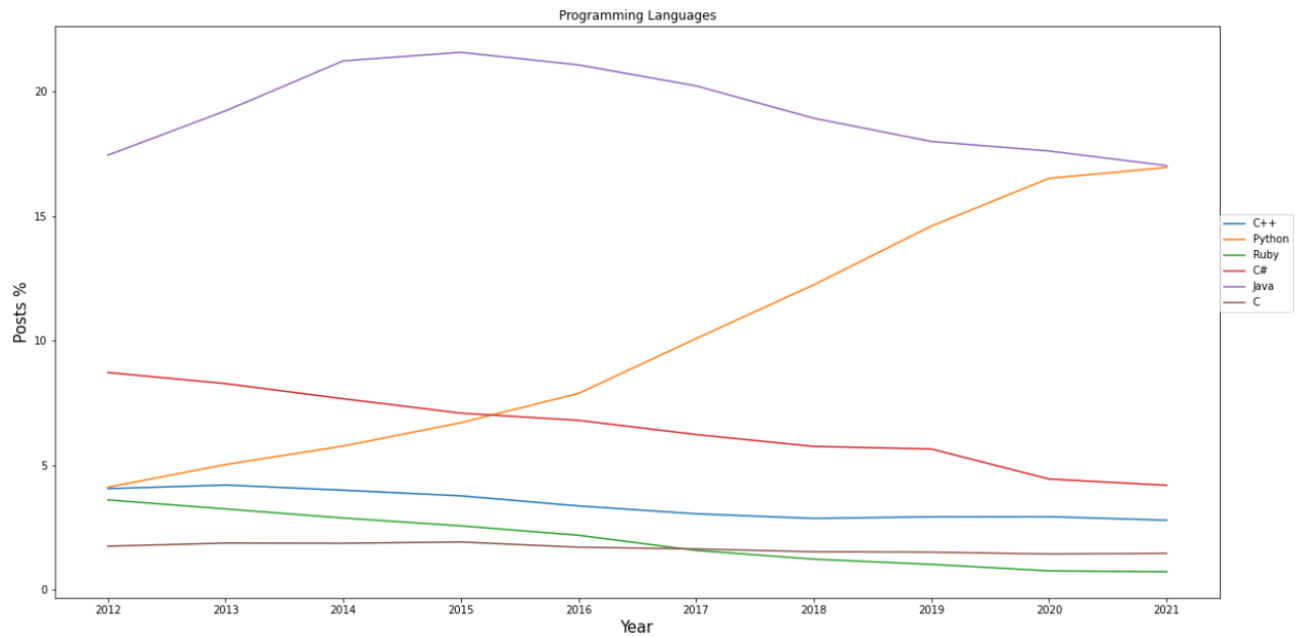
### Liczba postów z tagiem 'C' na przestrzeni lat 2012-2021

	year	posts
1.	2021	342 181 874 103
2.	2020	1 976 910 728 139
3.	2019	1 524 435 006 122
4.	2018	1 474 081 163 429
5.	2017	1 551 850 733 226
6.	2016	1 428 633 249 406
7.	2015	1 314 463 670 835
8.	2014	973 323 999 509
9.	2013	668 801 775 598
10.	2012	325 142 813 310

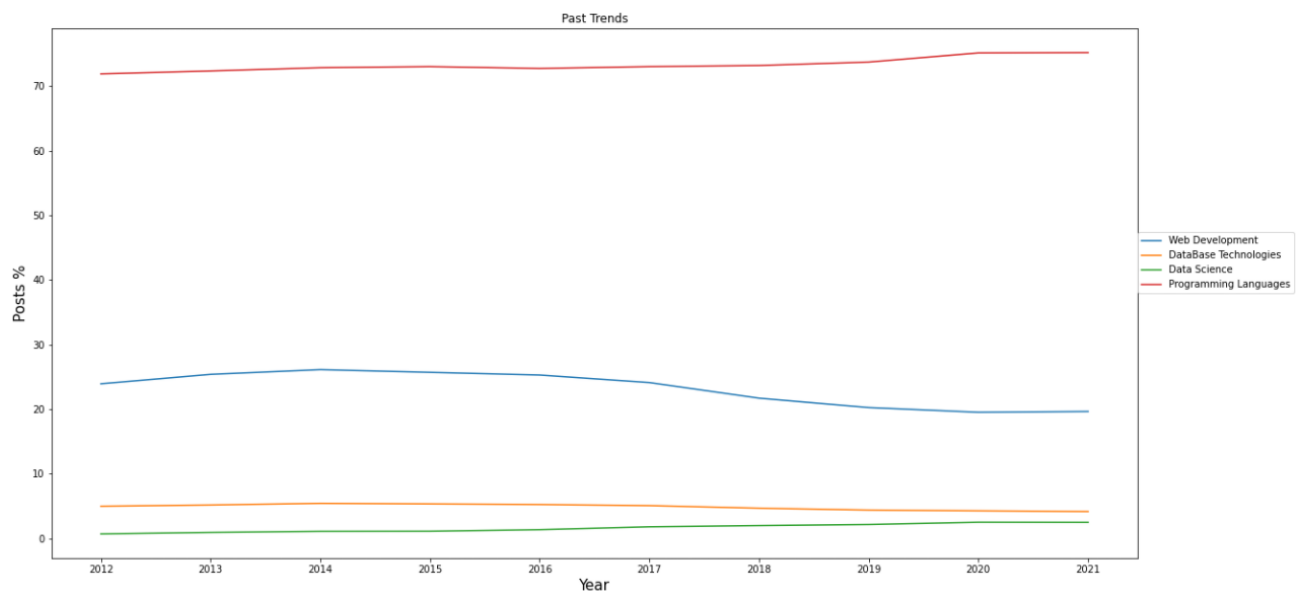
1 - 10 / 10 < >



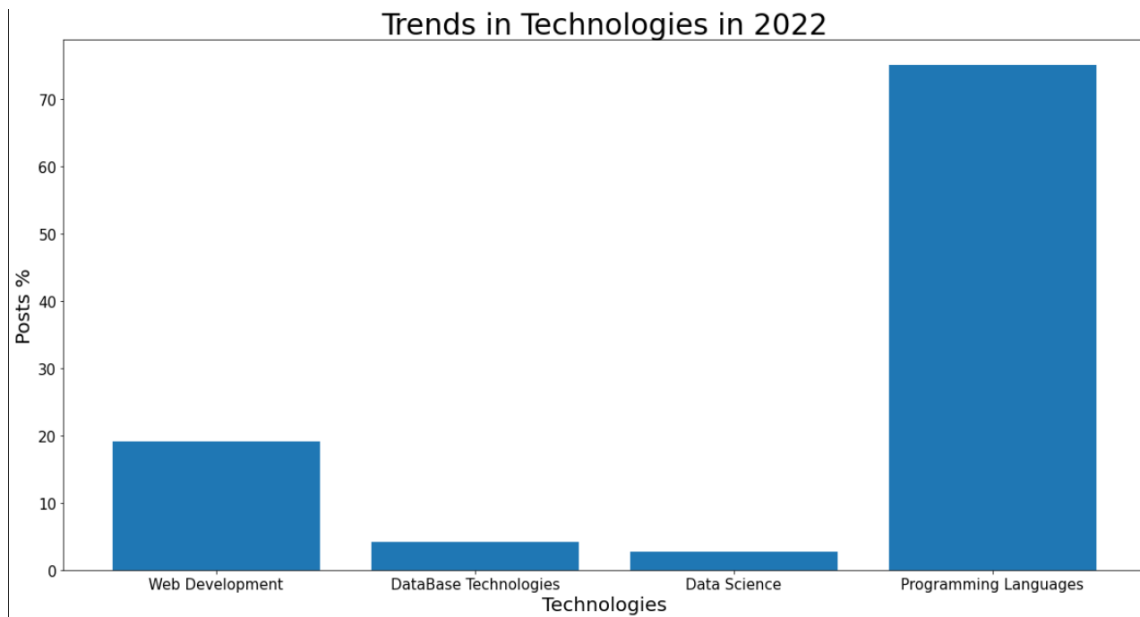
Wykres porównujący popularności różnych kategorii w ramach Programming Languages



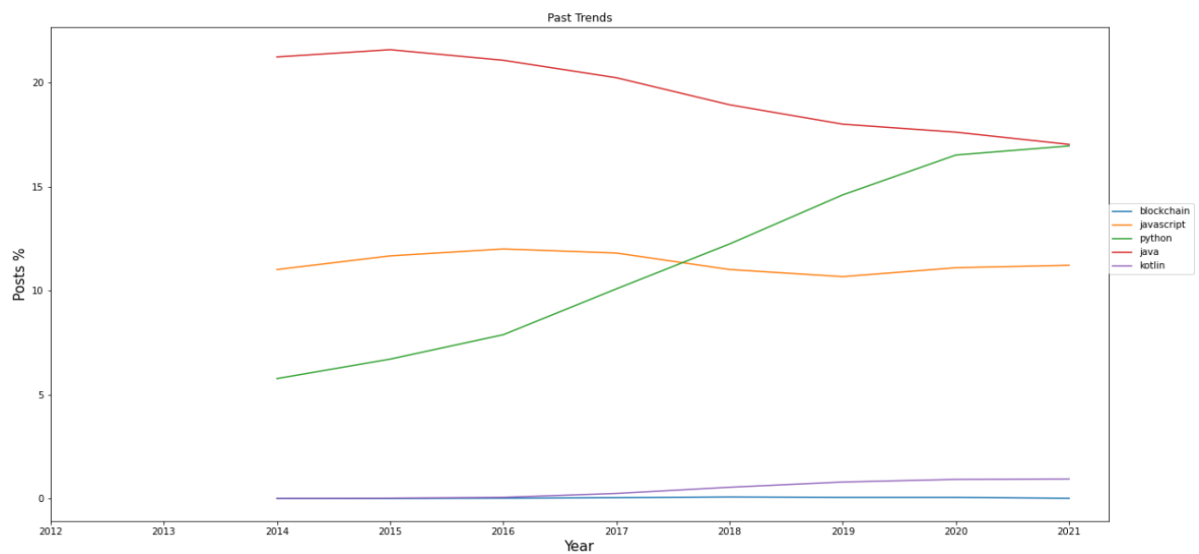
Wykres porównujący popularność przeszłych trendów w różnych kategoriach technologii



**Wykres porównujący popularność przyszłych trendów w różnych kategoriach technologii w 2022 roku**



**Wykres porównujący popularność przeszłych trendów od 2014 do 2021 roku**



Wykres porównujący popularność przyszłych trendów w 2032 roku

Trends in Technologies in 2032

