

# Remote Job Market Intelligence using Ethical Web Scraping

Presented by: Team G



# Project Overview

This project focuses on collecting real remote job listings from Remote OK, building an end-to-end data science pipeline. It emphasizes **ethical web scraping** to ensure compliance with web standards, paving the way for actionable insights into skills demand and job roles.



# Problem Statement & Challenges

---

## Key Questions

Addressing skills demand, job roles, job types, and geographic distribution in the remote job market.

## Project Challenges

Managing unstructured raw data, inefficiencies of manual analysis, and adhering to ethical web scraping constraints.

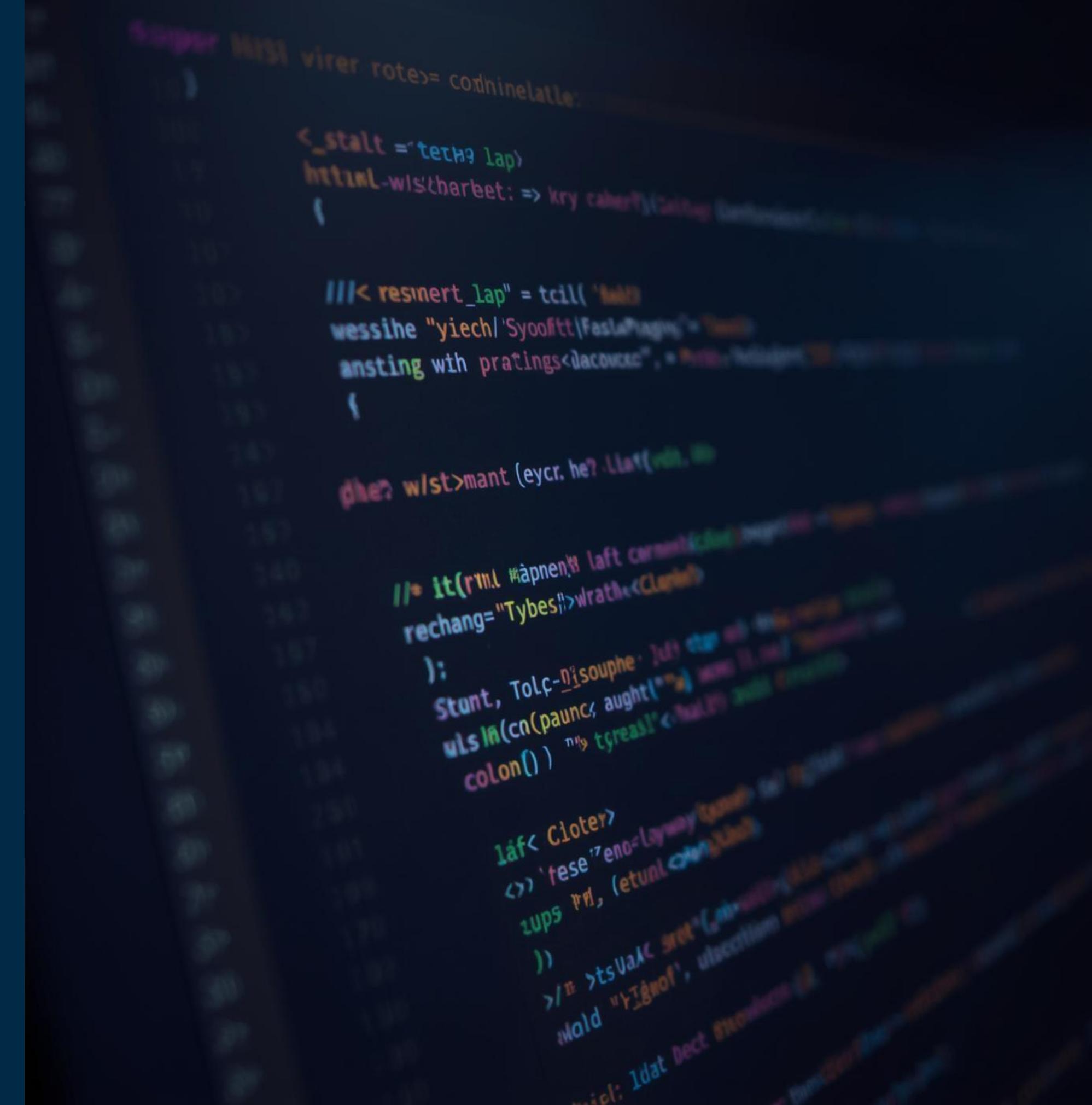
# Workflow Architecture

The project employs a comprehensive data science pipeline that includes stages such as website analysis, ethical web scraping, data cleaning, analysis, visualization, documentation, ethics review, and testing. Each stage is crucial and assigned to specific team members to ensure efficiency.



# Website Analysis

This section outlines the **website structure mapping** conducted by **Sarang**. It details the exploration of the Remote OK site, focusing on critical HTML elements such as Job Title, Company, Skills, Location, and Job Type to ensure efficient scraping and data extraction.



# Web Scraping Methodology

This section explains the ethical web scraping approach carried out by **Ansari and Rahul**. Using Selenium to handle dynamically rendered content, the team ensured compliance with web standards while efficiently gathering job listing data from Remote OK's platform.

A blurred screenshot of a computer monitor displaying a terminal window with code and numbers. The code appears to be in Python, dealing with file operations like reading and writing, and some numerical values. The background is dark, and the text is in various colors typical of a code editor.

```
84 13 1000
85 12 999
86 11 998
87 10 997
88 9 996
89 8 995
90 7 994
91 6 993
92 5 992
93 4 991
94 3 990
95 2 989
96 1 988
97 0 987
98 1 986
99 2 985
100 3 984
101 4 983
102 5 982
103 6 981
104 7 980
105 8 979
106 9 978
107 10 977
108 11 976
109 12 975
110 13 974
111 14 973
112 15 972
113 16 971
114 17 970
115 18 969
116 19 968
117 20 967
118 21 966
119 22 965
120 23 964
121 24 963
122 25 962
123 26 961
124 27 960
125 28 959
126 29 958
127 30 957
128 31 956
129 32 955
130 33 954
131 34 953
132 35 952
133 36 951
134 37 950
135 38 949
136 39 948
137 40 947
138 41 946
139 42 945
140 43 944
141 44 943
142 45 942
143 46 941
144 47 940
145 48 939
146 49 938
147 50 937
148 51 936
149 52 935
150 53 934
151 54 933
152 55 932
153 56 931
154 57 930
155 58 929
156 59 928
157 60 927
158 61 926
159 62 925
160 63 924
161 64 923
162 65 922
163 66 921
164 67 920
165 68 919
166 69 918
167 70 917
168 71 916
169 72 915
170 73 914
171 74 913
172 75 912
173 76 911
174 77 910
175 78 909
176 79 908
177 80 907
178 81 906
179 82 905
180 83 904
181 84 903
182 85 902
183 86 901
184 87 900
185 88 899
186 89 898
187 90 897
188 91 896
189 92 895
190 93 894
191 94 893
192 95 892
193 96 891
194 97 890
195 98 889
196 99 888
197 100 887
198 101 886
199 102 885
200 103 884
201 104 883
202 105 882
203 106 881
204 107 880
205 108 879
206 109 878
207 110 877
208 111 876
209 112 875
210 113 874
211 114 873
212 115 872
213 116 871
214 117 870
215 118 869
216 119 868
217 120 867
218 121 866
219 122 865
220 123 864
221 124 863
222 125 862
223 126 861
224 127 860
225 128 859
226 129 858
227 130 857
228 131 856
229 132 855
230 133 854
231 134 853
232 135 852
233 136 851
234 137 850
235 138 849
236 139 848
237 140 847
238 141 846
239 142 845
240 143 844
241 144 843
242 145 842
243 146 841
244 147 840
245 148 839
246 149 838
247 150 837
248 151 836
249 152 835
250 153 834
251 154 833
252 155 832
253 156 831
254 157 830
255 158 829
256 159 828
257 160 827
258 161 826
259 162 825
260 163 824
261 164 823
262 165 822
263 166 821
264 167 820
265 168 819
266 169 818
267 170 817
268 171 816
269 172 815
270 173 814
271 174 813
272 175 812
273 176 811
274 177 810
275 178 809
276 179 808
277 180 807
278 181 806
279 182 805
280 183 804
281 184 803
282 185 802
283 186 801
284 187 800
285 188 799
286 189 798
287 190 797
288 191 796
289 192 795
290 193 794
291 194 793
292 195 792
293 196 791
294 197 790
295 198 789
296 199 788
297 200 787
298 201 786
299 202 785
300 203 784
301 204 783
302 205 782
303 206 781
304 207 780
305 208 779
306 209 778
307 210 777
308 211 776
309 212 775
310 213 774
311 214 773
312 215 772
313 216 771
314 217 770
315 218 769
316 219 768
317 220 767
318 221 766
319 222 765
320 223 764
321 224 763
322 225 762
323 226 761
324 227 760
325 228 759
326 229 758
327 230 757
328 231 756
329 232 755
330 233 754
331 234 753
332 235 752
333 236 751
334 237 750
335 238 749
336 239 748
337 240 747
338 241 746
339 242 745
340 243 744
341 244 743
342 245 742
343 246 741
344 247 740
345 248 739
346 249 738
347 250 737
348 251 736
349 252 735
350 253 734
351 254 733
352 255 732
353 256 731
354 257 730
355 258 729
356 259 728
357 260 727
358 261 726
359 262 725
360 263 724
361 264 723
362 265 722
363 266 721
364 267 720
365 268 719
366 269 718
367 270 717
368 271 716
369 272 715
370 273 714
371 274 713
372 275 712
373 276 711
374 277 710
375 278 709
376 279 708
377 280 707
378 281 706
379 282 705
380 283 704
381 284 703
382 285 702
383 286 701
384 287 700
385 288 699
386 289 698
387 290 697
388 291 696
389 292 695
390 293 694
391 294 693
392 295 692
393 296 691
394 297 690
395 298 689
396 299 688
397 300 687
398 301 686
399 302 685
400 303 684
401 304 683
402 305 682
403 306 681
404 307 680
405 308 679
406 309 678
407 310 677
408 311 676
409 312 675
410 313 674
411 314 673
412 315 672
413 316 671
414 317 670
415 318 669
416 319 668
417 320 667
418 321 666
419 322 665
420 323 664
421 324 663
422 325 662
423 326 661
424 327 660
425 328 659
426 329 658
427 330 657
428 331 656
429 332 655
430 333 654
431 334 653
432 335 652
433 336 651
434 337 650
435 338 649
436 339 648
437 340 647
438 341 646
439 342 645
440 343 644
441 344 643
442 345 642
443 346 641
444 347 640
445 348 639
446 349 638
447 350 637
448 351 636
449 352 635
450 353 634
451 354 633
452 355 632
453 356 631
454 357 630
455 358 629
456 359 628
457 360 627
458 361 626
459 362 625
460 363 624
461 364 623
462 365 622
463 366 621
464 367 620
465 368 619
466 369 618
467 370 617
468 371 616
469 372 615
470 373 614
471 374 613
472 375 612
473 376 611
474 377 610
475 378 609
476 379 608
477 380 607
478 381 606
479 382 605
480 383 604
481 384 603
482 385 602
483 386 601
484 387 600
485 388 599
486 389 598
487 390 597
488 391 596
489 392 595
490 393 594
491 394 593
492 395 592
493 396 591
494 397 590
495 398 589
496 399 588
497 400 587
498 401 586
499 402 585
500 403 584
501 404 583
502 405 582
503 406 581
504 407 580
505 408 579
506 409 578
507 410 577
508 411 576
509 412 575
510 413 574
511 414 573
512 415 572
513 416 571
514 417 570
515 418 569
516 419 568
517 420 567
518 421 566
519 422 565
520 423 564
521 424 563
522 425 562
523 426 561
524 427 560
525 428 559
526 429 558
527 430 557
528 431 556
529 432 555
530 433 554
531 434 553
532 435 552
533 436 551
534 437 550
535 438 549
536 439 548
537 440 547
538 441 546
539 442 545
540 443 544
541 444 543
542 445 542
543 446 541
544 447 540
545 448 539
546 449 538
547 450 537
548 451 536
549 452 535
550 453 534
551 454 533
552 455 532
553 456 531
554 457 530
555 458 529
556 459 528
557 460 527
558 461 526
559 462 525
560 463 524
561 464 523
562 465 522
563 466 521
564 467 520
565 468 519
566 469 518
567 470 517
568 471 516
569 472 515
570 473 514
571 474 513
572 475 512
573 476 511
574 477 510
575 478 509
576 479 508
577 480 507
578 481 506
579 482 505
580 483 504
581 484 503
582 485 502
583 486 501
584 487 500
585 488 499
586 489 498
587 490 497
588 491 496
589 492 495
590 493 494
591 494 493
592 495 492
593 496 491
594 497 490
595 498 489
596 499 488
597 500 487
598 501 486
599 502 485
600 503 484
601 504 483
602 505 482
603 506 481
604 507 480
605 508 479
606 509 478
607 510 477
608 511 476
609 512 475
610 513 474
611 514 473
612 515 472
613 516 471
614 517 470
615 518 469
616 519 468
617 520 467
618 521 466
619 522 465
620 523 464
621 524 463
622 525 462
623 526 461
624 527 460
625 528 459
626 529 458
627 530 457
628 531 456
629 532 455
630 533 454
631 534 453
632 535 452
633 536 451
634 537 450
635 538 449
636 539 448
637 540 447
638 541 446
639 542 445
640 543 444
641 544 443
642 545 442
643 546 441
644 547 440
645 548 439
646 549 438
647 550 437
648 551 436
649 552 435
650 553 434
651 554 433
652 555 432
653 556 431
654 557 430
655 558 429
656 559 428
657 560 427
658 561 426
659 562 425
660 563 424
661 564 423
662 565 422
663 566 421
664 567 420
665 568 419
666 569 418
667 570 417
668 571 416
669 572 415
670 573 414
671 574 413
672 575 412
673 576 411
674 577 410
675 578 409
676 579 408
677 580 407
678 581 406
679 582 405
680 583 404
681 584 403
682 585 402
683 586 401
684 587 400
685 588 399
686 589 398
687 590 397
688 591 396
689 592 395
690 593 394
691 594 393
692 595 392
693 596 391
694 597 390
695 598 389
696 599 388
697 600 387
698 601 386
699 602 385
700 603 384
701 604 383
702 605 382
703 606 381
704 607 380
705 608 379
706 609 378
707 610 377
708 611 376
709 612 375
710 613 374
711 614 373
712 615 372
713 616 371
714 617 370
715 618 369
716 619 368
717 620 367
718 621 366
719 622 365
720 623 364
721 624 363
722 625 362
723 6
```

# Data Cleaning

In this phase, we focused on refining the raw data collected from web scraping. Key tasks included **removing duplicates**, addressing missing values, and ensuring text normalization to create a clean, structured dataset that facilitates accurate analysis and insights generation. Done by

**Tejas**



# Data Analysis Insights

In this section, which is conducted by **Himanush**, we present the **key findings** from our data analysis, focusing on trends in job roles, skills demand, and geographical distribution. These actionable insights provide a comprehensive understanding of the remote job market landscape.



# Data Visualization and Dashboard Creation

This section was handled by **Huda Saiyed** and focuses on transforming the analyzed data into clear and meaningful **visual representations**. It includes the creation of charts and visual summaries, such as skill demand, job type distribution, job trends over time, and top hiring locations. These visualizations effectively communicate key insights and support data-driven decision-making.



# Documentation

This section was handled by **Priyanga** and focuses on compiling all source code contributions from the team. The collected scripts were carefully reviewed, refined, and integrated into a single, well-organized Jupyter Notebook (Team\_G.ipynb). The **documentation** dearly presents the complete project workflow in a structured manner, making it easy to understand, review, and present.

```
226     ffflush(stdin);
227     scanf("%[^\\n]",R.note);
228     if(fwrite(&R,sizeof(R),1,fp)){
229         gotoxy(5,12);
230         puts("Note is saved sucessfully");
231         fclose(fp);
232     }else{
233         gotoxy(5,12);
234         SetColor(12);
235         puts("\aFail to save!!\a");
236         ClearColor();
237         gotoxy(5,15);
238         printf("Press any key.....");
239         getch();
240         fclose(fp);
```

# Testing & Preparation

This section outlines the final **validation process** led by **Prajakta**, ensuring the pipeline's integrity. It also highlights the collaborative efforts of **Shifa and Vimalesh** in organizing the **presentation** flow, emphasizing key insights and ethical considerations for effective communication.



# THANK YOU

