



# Retail Sales Predictions: Machine Learning Project

Welcome to our comprehensive project report on retail sales predictions using machine learning. Our team of 12 students from B.Tech CSE 2nd Year, Section A, led by Prerna Sharma, worked under the mentorship of Manab Das Sir to develop predictive models for weekly retail sales.

This presentation will walk you through our methodology, data analysis, model building process, and key findings that demonstrate how machine learning can be applied to real-world retail scenarios for effective sales forecasting.

## Team Members

- ❖ RINKU (Team leader) (BCS2023208)
- ❖ SHYOG
- ❖ JAYESH PRAKASH SHUKLA
- ❖ ADITYA RAJ
- ❖ PRATHAM KUMAR
- ❖ SUPRIYA MISHRA
- ❖ ASHISH KUMAR
- ❖ SHEKHAR KANAUJIYA
- ❖ POORNIMA GUPTA
- ❖ PALLAVI GUPTA
- ❖ HIMANSHU GANGWAR
- ❖ SAURABH KUMAR SINGH



# Project Overview & Objectives



## Primary Goal

Predict weekly sales of retail stores using machine learning models to assist in inventory planning and business forecasting



## Data Sources

Merged multiple datasets (features.csv, stores.csv, train.csv) to create a comprehensive analysis foundation



## Approach

Implemented data cleaning, exploratory analysis, feature engineering, and multiple regression models to identify the most effective prediction method

# Technical Framework & Libraries



## NumPy & Pandas

For numerical operations, array handling, data manipulation and analysis using DataFrames



## Matplotlib & Seaborn

For basic plotting, advanced visualizations including boxplots and heatmaps



## Scikit-learn

For preprocessing, model building (Linear Regression, KNN, Random Forest), and evaluation metrics



## Statsmodels

For calculating Variance Inflation Factor (VIF) to detect multicollinearity



```
Positie 6: Geen resultaat voor
  1: 1
  2: 2
  3: 3
  4: 4
  5: 5
  6: 6
  7: 7
  8: 8
  9: 9
  10: 10
  11: 11
  12: 12
  13: 13
  14: 14
  15: 15
  16: 16
  17: 17
  18: 18
  19: 19
  20: 20
  21: 21
  22: 22
  23: 23
  24: 24
  25: 25
  26: 26
  27: 27
  28: 28
  29: 29
  30: 30
  31: 31
  32: 32
  33: 33
  34: 34
  35: 35
  36: 36
  37: 37
  38: 38
  39: 39
  40: 40
  41: 41
  42: 42
  43: 43
  44: 44
  45: 45
  46: 46
  47: 47
  48: 48
  49: 49
  50: 50
  51: 51
  52: 52
  53: 53
  54: 54
  55: 55
  56: 56
  57: 57
  58: 58
  59: 59
  60: 60
  61: 61
  62: 62
  63: 63
  64: 64
  65: 65
  66: 66
  67: 67
  68: 68
  69: 69
  70: 70
  71: 71
  72: 72
  73: 73
  74: 74
  75: 75
  76: 76
  77: 77
  78: 78
  79: 79
  80: 80
  81: 81
  82: 82
  83: 83
  84: 84
  85: 85
  86: 86
  87: 87
  88: 88
  89: 89
  90: 90
  91: 91
  92: 92
  93: 93
  94: 94
  95: 95
  96: 96
  97: 97
  98: 98
  99: 99
  100: 100
  101: 101
  102: 102
  103: 103
  104: 104
  105: 105
  106: 106
  107: 107
  108: 108
  109: 109
  110: 110
  111: 111
  112: 112
  113: 113
  114: 114
  115: 115
  116: 116
  117: 117
  118: 118
  119: 119
  120: 120
  121: 121
  122: 122
  123: 123
  124: 124
  125: 125
  126: 126
  127: 127
  128: 128
  129: 129
  130: 130
  131: 131
  132: 132
  133: 133
  134: 134
  135: 135
  136: 136
  137: 137
  138: 138
  139: 139
  140: 140
  141: 141
  142: 142
  143: 143
  144: 144
  145: 145
  146: 146
  147: 147
  148: 148
  149: 149
  150: 150
  151: 151
  152: 152
  153: 153
  154: 154
  155: 155
  156: 156
  157: 157
  158: 158
  159: 159
  160: 160
  161: 161
  162: 162
  163: 163
  164: 164
  165: 165
  166: 166
  167: 167
  168: 168
  169: 169
  170: 170
  171: 171
  172: 172
  173: 173
  174: 174
  175: 175
  176: 176
  177: 177
  178: 178
  179: 179
  180: 180
  181: 181
  182: 182
  183: 183
  184: 184
  185: 185
  186: 186
  187: 187
  188: 188
  189: 189
  190: 190
  191: 191
  192: 192
  193: 193
  194: 194
  195: 195
  196: 196
  197: 197
  198: 198
  199: 199
  200: 200
  201: 201
  202: 202
  203: 203
  204: 204
  205: 205
  206: 206
  207: 207
  208: 208
  209: 209
  210: 210
  211: 211
  212: 212
  213: 213
  214: 214
  215: 215
  216: 216
  217: 217
  218: 218
  219: 219
  220: 220
  221: 221
  222: 222
  223: 223
  224: 224
  225: 225
  226: 226
  227: 227
  228: 228
  229: 229
  230: 230
  231: 231
  232: 232
  233: 233
  234: 234
  235: 235
  236: 236
  237: 237
  238: 238
  239: 239
  240: 240
  241: 241
  242: 242
  243: 243
  244: 244
  245: 245
  246: 246
  247: 247
  248: 248
  249: 249
  250: 250
  251: 251
  252: 252
  253: 253
  254: 254
  255: 255
  256: 256
  257: 257
  258: 258
  259: 259
  260: 260
  261: 261
  262: 262
  263: 263
  264: 264
  265: 265
  266: 266
  267: 267
  268: 268
  269: 269
  270: 270
  271: 271
  272: 272
  273: 273
  274: 274
  275: 275
  276: 276
  277: 277
  278: 278
  279: 279
  280: 280
  281: 281
  282: 282
  283: 283
  284: 284
  285: 285
  286: 286
  287: 287
  288: 288
  289: 289
  290: 290
  291: 291
  292: 292
  293: 293
  294: 294
  295: 295
  296: 296
  297: 297
  298: 298
  299: 299
  300: 300
  301: 301
  302: 302
  303: 303
  304: 304
  305: 305
  306: 306
  307: 307
  308: 308
  309: 309
  310: 310
  311: 311
  312: 312
  313: 313
  314: 314
  315: 315
  316: 316
  317: 317
  318: 318
  319: 319
  320: 320
  321: 321
  322: 322
  323: 323
  324: 324
  325: 325
  326: 326
  327: 327
  328: 328
  329: 329
  330: 330
  331: 331
  332: 332
  333: 333
  334: 334
  335: 335
  336: 336
  337: 337
  338: 338
  339: 339
  340: 340
  341: 341
  342: 342
  343: 343
  344: 344
  345: 345
  346: 346
  347: 347
  348: 348
  349: 349
  350: 350
  351: 351
  352: 352
  353: 353
  354: 354
  355: 355
  356: 356
  357: 357
  358: 358
  359: 359
  360: 360
  361: 361
  362: 362
  363: 363
  364: 364
  365: 365
  366: 366
  367: 367
  368: 368
  369: 369
  370: 370
  371: 371
  372: 372
  373: 373
  374: 374
  375: 375
  376: 376
  377: 377
  378: 378
  379: 379
  380: 380
  381: 381
  382: 382
  383: 383
  384: 384
  385: 385
  386: 386
  387: 387
  388: 388
  389: 389
  390: 390
  391: 391
  392: 392
  393: 393
  394: 394
  395: 395
  396: 396
  397: 397
  398: 398
  399: 399
  400: 400
  401: 401
  402: 402
  403: 403
  404: 404
  405: 405
  406: 406
  407: 407
  408: 408
  409: 409
  410: 410
  411: 411
  412: 412
  413: 413
  414: 414
  415: 415
  416: 416
  417: 417
  418: 418
  419: 419
  420: 420
  421: 421
  422: 422
  423: 423
  424: 424
  425: 425
  426: 426
  427: 427
  428: 428
  429: 429
  430: 430
  431: 431
  432: 432
  433: 433
  434: 434
  435: 435
  436: 436
  437: 437
  438: 438
  439: 439
  440: 440
  441: 441
  442: 442
  443: 443
  444: 444
  445: 445
  446: 446
  447: 447
  448: 448
  449: 449
  450: 450
  451: 451
  452: 452
  453: 453
  454: 454
  455: 455
  456: 456
  457: 457
  458: 458
  459: 459
  460: 460
  461: 461
  462: 462
  463: 463
  464: 464
  465: 465
  466: 466
  467: 467
  468: 468
  469: 469
  470: 470
  471: 471
  472: 472
  473: 473
  474: 474
  475: 475
  476: 476
  477: 477
  478: 478
  479: 479
  480: 480
  481: 481
  482: 482
  483: 483
  484: 484
  485: 485
  486: 486
  487: 487
  488: 488
  489: 489
  490: 490
  491: 491
  492: 492
  493: 493
  494: 494
  495: 495
  496: 496
  497: 497
  498: 498
  499: 499
  500: 500
  501: 501
  502: 502
  503: 503
  504: 504
  505: 505
  506: 506
  507: 507
  508: 508
  509: 509
  510: 510
  511: 511
  512: 512
  513: 513
  514: 514
  515: 515
  516: 516
  517: 517
  518: 518
  519: 519
  520: 520
  521: 521
  522: 522
  523: 523
  524: 524
  525: 525
  526: 526
  527: 527
  528: 528
  529: 529
  530: 530
  531: 531
  532: 532
  533: 533
  534: 534
  535: 535
  536: 536
  537: 537
  538: 538
  539: 539
  540: 540
  541: 541
  542: 542
  543: 543
  544: 544
  545: 545
  546: 546
  547: 547
  548: 548
  549: 549
  550: 550
  551: 551
  552: 552
  553: 553
  554: 554
  555: 555
  556: 556
  557: 557
  558: 558
  559: 559
  560: 560
  561: 561
  562: 562
  563: 563
  564: 564
  565: 565
  566: 566
  567: 567
  568: 568
  569: 569
  570: 570
  571: 571
  572: 572
  573: 573
  574: 574
  575: 575
  576: 576
  577: 577
  578: 578
  579: 579
  580: 580
  581: 581
  582: 582
  583: 583
  584: 584
  585: 585
  586: 586
  587: 587
  588: 588
  589: 589
  590: 590
  591: 591
  592: 592
  593: 593
  594: 594
  595: 595
  596: 596
  597: 597
  598: 598
  599: 599
  600: 600
  601: 601
  602: 602
  603: 603
  604: 604
  605: 605
  606: 606
  607: 607
  608: 608
  609: 609
  610: 610
  611: 611
  612: 612
  613: 613
  614: 614
  615: 615
  616: 616
  617: 617
  618: 618
  619: 619
  620: 620
  621: 621
  622: 622
  623: 623
  624: 624
  625: 625
  626: 626
  627: 627
  628: 628
  629: 629
  630: 630
  631: 631
  632: 632
  633: 633
  634: 634
  635: 635
  636: 636
  637: 637
  638: 638
  639: 639
  640: 640
  641: 641
  642: 642
  643: 643
  644: 644
  645: 645
  646: 646
  647: 647
  648: 648
  649: 649
  650: 650
  651: 651
  652: 652
  653: 653
  654: 654
  655: 655
  656: 656
  657: 657
  658: 658
  659: 659
  660: 660
  661: 661
  662: 662
  663: 663
  664: 664
  665: 665
  666: 666
  667: 667
  668: 668
  669: 669
  670: 670
  671: 671
  672: 672
  673: 673
  674: 674
  675: 675
  676: 676
  677: 677
  678: 678
  679: 679
  680: 680
  681: 681
  682: 682
  683: 683
  684: 684
  685: 685
  686: 686
  687: 687
  688: 688
  689: 689
  690: 690
  691: 691
  692: 692
  693: 693
  694: 694
  695: 695
  696: 696
  697: 697
  698: 698
  699: 699
  700: 700
  701: 701
  702: 702
  703: 703
  704: 704
  705: 705
  706: 706
  707: 707
  708: 708
  709: 709
  710: 710
  711: 711
  712: 712
  713: 713
  714: 714
  715: 715
  716: 716
  717: 717
  718: 718
  719: 719
  720: 720
  721: 721
  722: 722
  723: 723
  724: 724
  725: 725
  726: 726
  727: 727
  728: 728
  729: 729
  730: 730
  731: 731
  732: 732
  733: 733
  734: 734
  735: 735
  736: 736
  737: 737
  738: 738
  739: 739
  740: 740
  741: 741
  742: 742
  743: 743
  744: 744
  745: 745
  746: 746
  747: 747
  748: 748
  749: 749
  750: 750
  751: 751
  752: 752
  753: 753
  754: 754
  755: 755
  756: 756
  757: 757
  758: 758
  759: 759
  760: 760
  761: 761
  762: 762
  763: 763
  764: 764
  765: 765
  766: 766
  767: 767
  768: 768
  769: 769
  770: 770
  771: 771
  772: 772
  773: 773
  774: 774
```

# Data Preparation Process

# Data Loading & Merging

Loaded features.csv, stores.csv, and train.csv using Pandas and merged them on Store and Date columns to create a unified dataset for analysis

# Data Cleaning

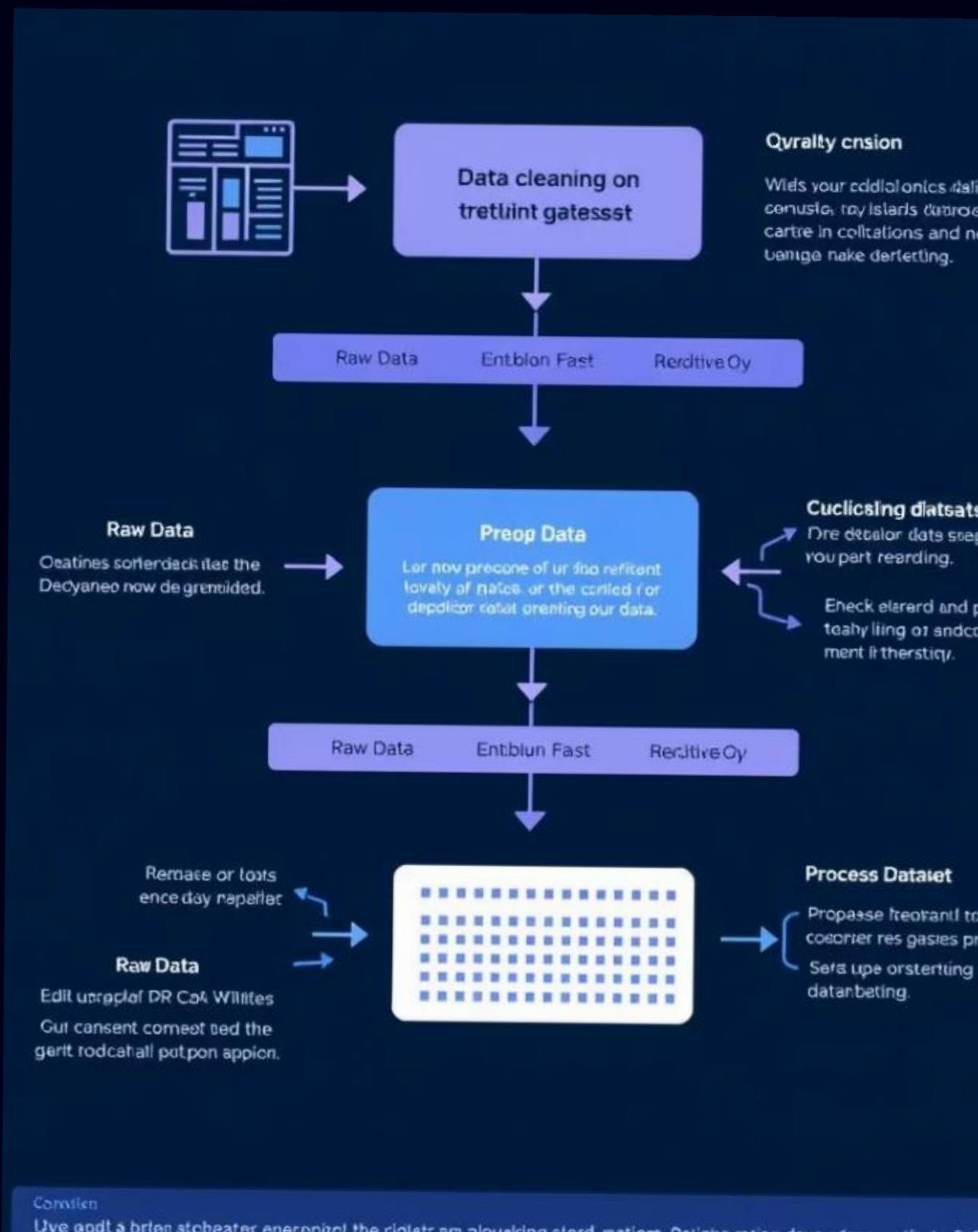
Removed unnecessary columns (Dept, Date, MarkDown1-5) to reduce noise and eliminate duplicate columns like IsHoliday\_y for consistency

## Feature Engineering

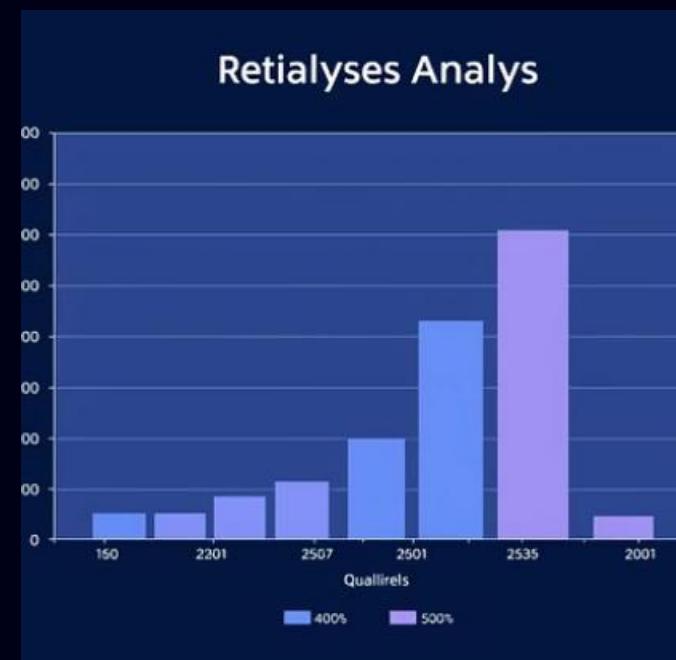
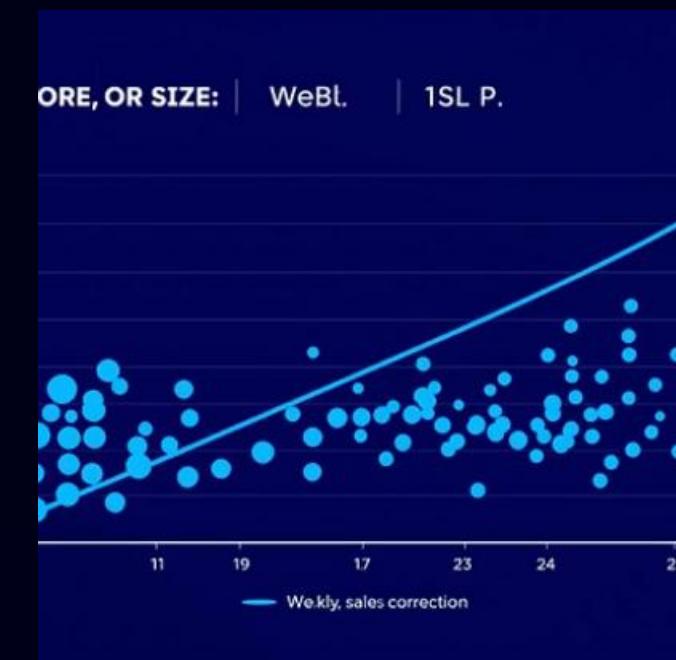
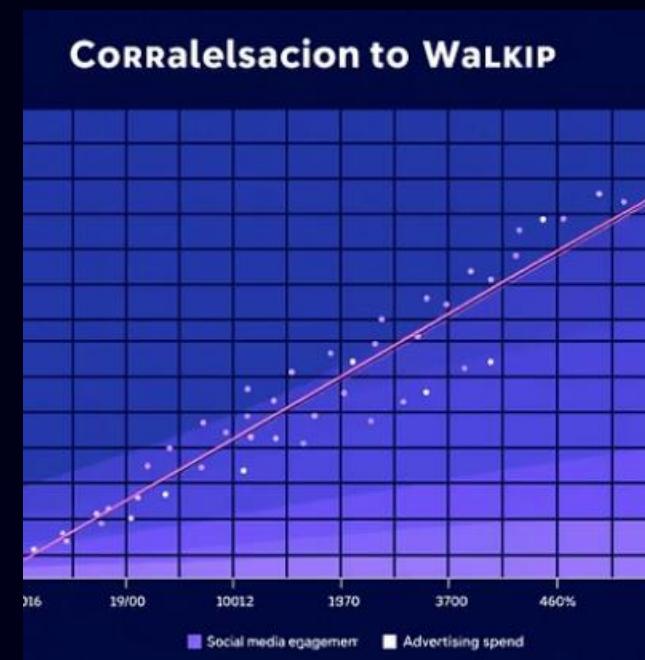
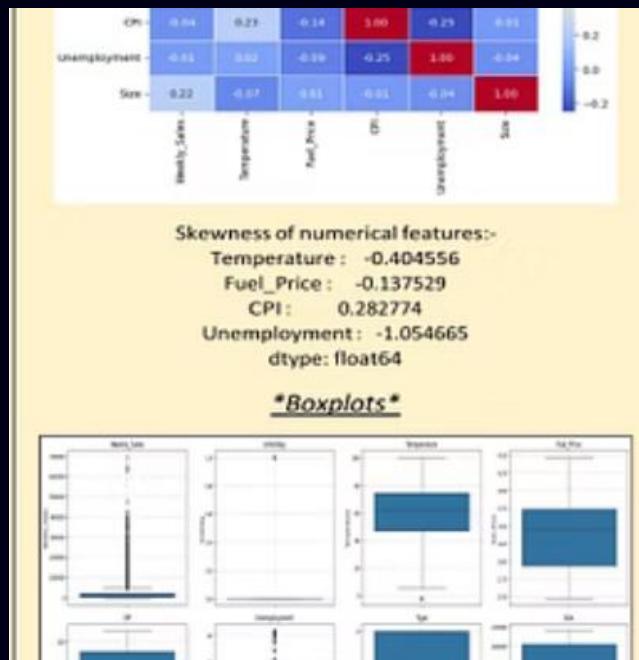
Performed label encoding for categorical variables and checked for multicollinearity using Variance Inflation Factor (VIF) to improve model performance

## Outlier Handling

Detected outliers using the IQR method and removed them to enhance model accuracy and reliability



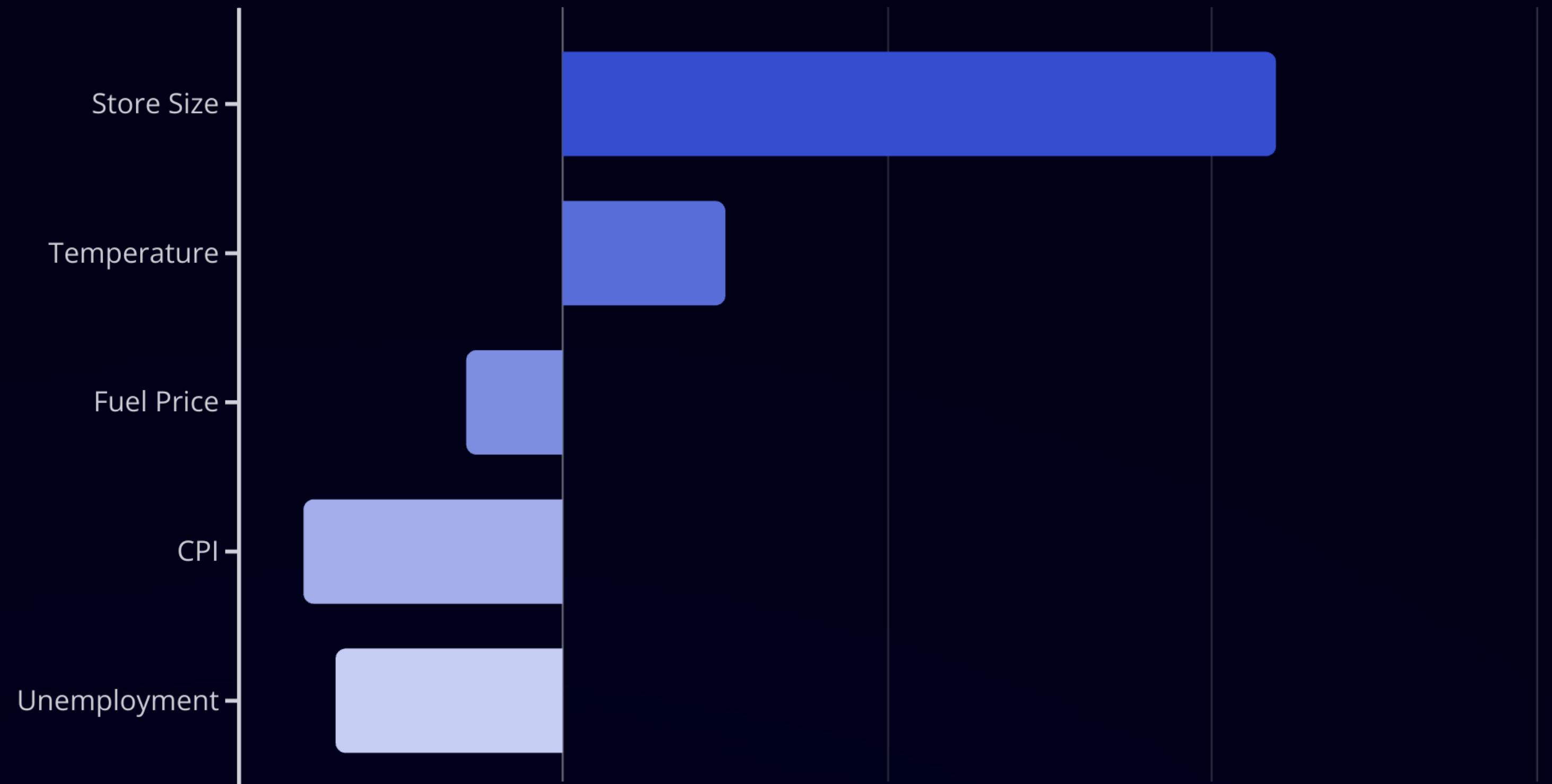
# Exploratory Data Analysis



Our exploratory data analysis revealed important relationships between variables. We used boxplots to detect outliers in numeric features and created scatter plots to visualize relationships between Weekly\_Sales and factors like Temperature, CPI, Unemployment, and Fuel\_Price.

The correlation analysis showed that store Size had the highest positive correlation with Weekly\_Sales (0.22), while CPI and Unemployment showed slight negative correlations, providing valuable insights for our modeling approach.

# Key Correlation Findings





## Model Building & Evaluation

3

0.0619

10855.12

### Models Tested

Linear Regression, K-Nearest Neighbors, Random  
Forest Regressor

### Best R<sup>2</sup> Score

Achieved by Random Forest Regressor

### Lowest RMSE

Random Forest outperformed other models

We implemented three different regression models to predict weekly sales. The dataset was split into training and testing sets to evaluate model performance. Each model was trained on the same data and evaluated using standard metrics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R<sup>2</sup> Score.

While all models showed modest predictive power, the Random Forest Regressor demonstrated the best performance across all metrics, indicating its superior ability to capture the complex relationships in retail sales data.

# Model Performance Comparison

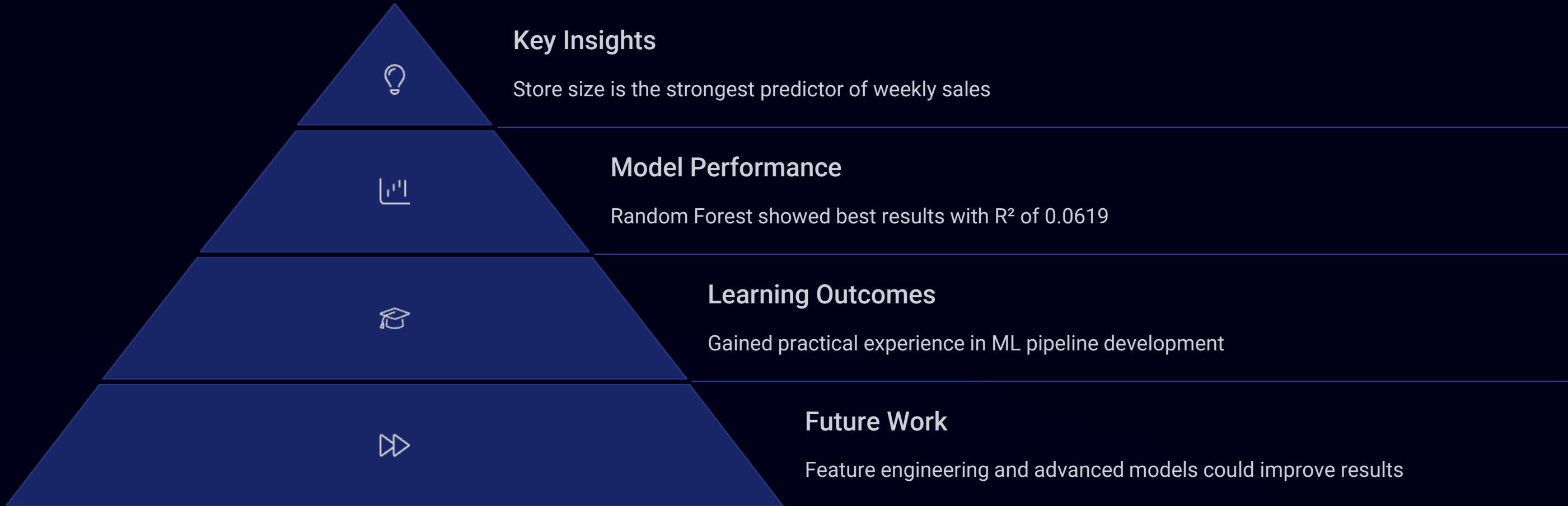


Model	RMSE	MAE	R <sup>2</sup> Score
Linear Regression	10914.82	8563.94	0.0516
K-Nearest Neighbors	11508.30	8664.09	-0.0544
Random Forest Regressor	10855.12	8322.19	0.0619

Our comprehensive model evaluation revealed that the Random Forest Regressor outperformed both Linear Regression and K-Nearest Neighbors across all metrics. It achieved the lowest RMSE (10855.12) and MAE (8322.19), indicating better prediction accuracy.

The R<sup>2</sup> score of 0.0619, while modest, was the highest among the three models. This suggests that Random Forest was able to capture more of the variance in weekly sales data, likely due to its ability to model non-linear relationships and handle feature interactions more effectively.

# Conclusions & Future Directions



This project provided valuable hands-on experience in applying machine learning to retail sales prediction. While our models showed modest predictive power, they demonstrated the potential of data-driven approaches in retail forecasting. The Random Forest model's superior performance highlights the importance of capturing non-linear relationships in sales data.

Future work could focus on more sophisticated feature engineering, incorporating temporal patterns, and exploring ensemble methods to further improve prediction accuracy. We extend our gratitude to our mentor Hitesh Panwar for his guidance and to all team members for their dedication throughout this learning journey.

# Thank You