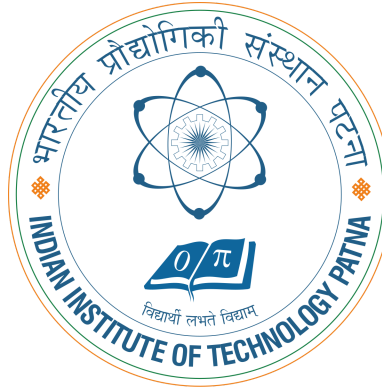


CS571: Artificial Intelligence Lab

Indian Institute of Technology Patna



ASSIGNMENT 7

Linear Regression

Group Members

- Kartik Kailas Mouli (2001CS35)
- Rohit Ranjan (2001CS56)
- Siddhant Kumar (2001CS70)

Step 1: Processing dataset

| SNo | | Newspaper | Daily | Sunday |
|-----|----|----------------------------|----------|----------|
| 0 | 0 | Baltimore Sun | 391.952 | 488.506 |
| 1 | 1 | Boston Globe | 516.981 | 798.298 |
| 2 | 2 | Boston Herald | 355.628 | 235.084 |
| 3 | 3 | Charlotte Observer | 238.555 | 299.451 |
| 4 | 4 | Chicago Sun Times | 537.780 | 559.093 |
| 5 | 5 | Chicago Tribune | 733.775 | 1133.249 |
| 6 | 6 | Cincinnati Enquirer | 198.832 | 348.744 |
| 7 | 7 | Denver Post | 252.624 | 417.779 |
| 8 | 8 | Des Moines Register | 206.204 | 344.522 |
| 9 | 9 | Hartford Courant | 231.177 | 323.084 |
| 10 | 10 | Houston Chronicle | 449.755 | 620.752 |
| 11 | 11 | Kansas City Star | 288.571 | 423.305 |
| 12 | 12 | Los Angeles Daily News | 185.736 | 202.614 |
| 13 | 13 | Los Angeles Times | 1164.388 | 1531.527 |
| 14 | 14 | Miami Herald | 444.581 | 553.479 |
| 15 | 15 | Minneapolis Star Tribune | 412.871 | 685.975 |
| 16 | 16 | New Orleans Times-Picayune | 272.280 | 324.241 |
| 17 | 17 | New York Daily News | 781.796 | 983.240 |
| 18 | 18 | New York Times | 1209.225 | 1762.015 |
| 19 | 19 | Newsday | 825.512 | 960.308 |
| 20 | 20 | Omaha World Herald | 223.748 | 284.611 |
| 21 | 21 | Orange County Register | 354.843 | 407.760 |
| 22 | 22 | Philadelphia Inquirer | 515.523 | 982.663 |
| 23 | 23 | Pittsburgh Press | 220.465 | 557.000 |
| 24 | 24 | Portland Oregonian | 337.672 | 440.923 |

| | | | | |
|----|----|--------------------------------|---------|----------|
| 25 | 25 | Providence Journal-Bulletin | 197.120 | 268.060 |
| 26 | 26 | Rochester Democrat & Chronicle | 133.239 | 262.048 |
| 27 | 27 | Rocky Mountain News | 374.009 | 432.502 |
| 28 | 28 | Sacramento Bee | 273.844 | 338.355 |
| 29 | 29 | San Francisco Chronicle | 570.364 | 704.322 |
| 30 | 30 | St. Louis Post-Dispatch | 391.286 | 585.681 |
| 31 | 31 | St. Paul Pioneer Press | 201.860 | 267.781 |
| 32 | 32 | Tampa Tribune | 321.626 | 408.343 |
| 33 | 33 | Washington Post | 838.902 | 1165.567 |

Step 2: Data Analysis

Min Daily Sales: 133.239

Max Daily Sales: 1209.225

Avg Daily Sales: 430.9624705882353

Min Daily Sales Threshold: 173.2107

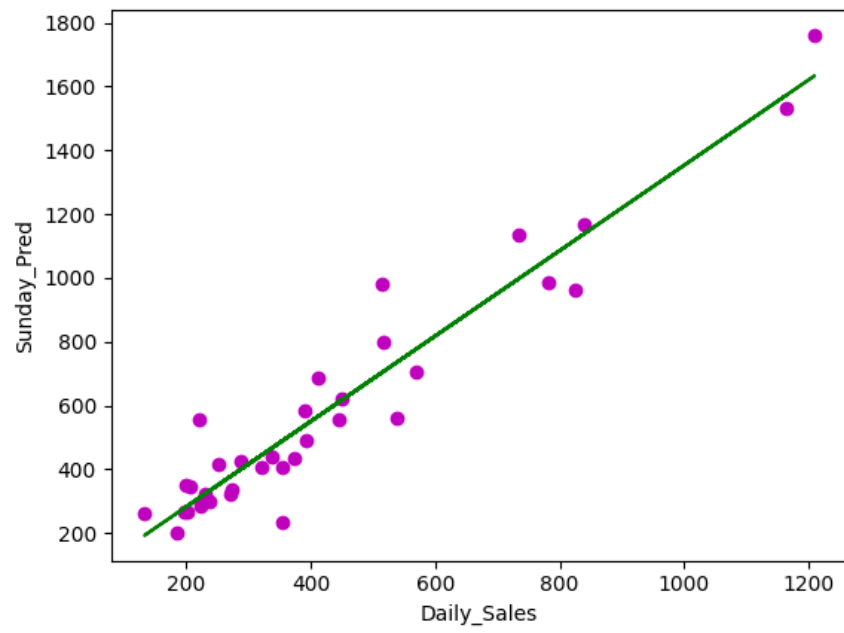
Max Daily Sales Threshold: 1571.9924999999998

Avg Daily Sales Threshold: 560.2512117647059

Step 3: Estimate coefficient

[13.83562990963594, 1.3397147576841262]

Step 4: Plot Graph



Step 5: Check which newspaper should stop Sunday circulation

| SNo | | Newspaper | Daily | Sunday | Sunday Pred | Stop_Sunday_Min | Stop_Sunday_Max | Stop_Sunday_Avg |
|-----|----|------------------------|---------|----------|-------------|-----------------|-----------------|-----------------|
| 0 | 0 | Baltimore Sun | 391.952 | 488.506 | 538.939509 | Yes | No | No |
| 1 | 1 | Boston Globe | 516.981 | 798.298 | 706.442705 | Yes | No | Yes |
| 2 | 2 | Boston Herald | 355.628 | 235.084 | 490.275710 | Yes | No | No |
| 3 | 3 | Charlotte Observer | 238.555 | 299.451 | 333.431284 | Yes | No | No |
| 4 | 4 | Chicago Sun Times | 537.780 | 559.093 | 734.307432 | Yes | No | Yes |
| 5 | 5 | Chicago Tribune | 733.775 | 1133.249 | 996.884826 | Yes | No | Yes |
| 6 | 6 | Cincinnati Enquirer | 198.832 | 348.744 | 280.213795 | Yes | No | No |
| 7 | 7 | Denver Post | 252.624 | 417.779 | 352.279731 | Yes | No | No |
| 8 | 8 | Des Moines Register | 206.204 | 344.522 | 290.090172 | Yes | No | No |
| 9 | 9 | Hartford Courant | 231.177 | 323.084 | 323.546868 | Yes | No | No |
| 10 | 10 | Houston Chronicle | 449.755 | 620.752 | 616.379041 | Yes | No | Yes |
| 11 | 11 | Kansas City Star | 288.571 | 423.305 | 400.438457 | Yes | No | No |
| 12 | 12 | Los Angeles Daily News | 185.736 | 202.614 | 262.668890 | Yes | No | No |

| | | | | | | | | |
|----|----|--------------------------------|----------|----------|-------------|-----|-----|-----|
| 13 | 13 | Los Angeles Times | 1164.388 | 1531.527 | 1573.783417 | Yes | Yes | Yes |
| 14 | 14 | Miami Herald | 444.581 | 553.479 | 609.447357 | Yes | No | Yes |
| 15 | 15 | Minneapolis Star Tribune | 412.871 | 685.975 | 566.965002 | Yes | No | Yes |
| 16 | 16 | New Orleans Times-Picayune | 272.280 | 324.241 | 378.613164 | Yes | No | No |
| 17 | 17 | New York Daily News | 781.796 | 983.240 | 1061.219269 | Yes | No | Yes |
| 18 | 18 | New York Times | 1209.225 | 1762.015 | 1633.852208 | Yes | Yes | Yes |
| 19 | 19 | Newsday | 825.512 | 960.308 | 1119.786239 | Yes | No | Yes |
| 20 | 20 | Omaha World Herald | 223.748 | 284.611 | 313.594128 | Yes | No | No |
| 21 | 21 | Orange County Register | 354.843 | 407.760 | 489.224034 | Yes | No | No |
| 22 | 22 | Philadelphia Inquirer | 515.523 | 982.663 | 704.489401 | Yes | No | Yes |
| 23 | 23 | Pittsburgh Press | 220.465 | 557.000 | 309.195844 | Yes | No | No |
| 24 | 24 | Portland Oregonian | 337.672 | 440.923 | 466.219792 | Yes | No | No |
| 25 | 25 | Providence Journal-Bulletin | 197.120 | 268.060 | 277.920203 | Yes | No | No |
| 26 | 26 | Rochester Democrat & Chronicle | 133.239 | 262.048 | 192.337885 | Yes | No | No |
| 27 | 27 | Rocky Mountain News | 374.009 | 432.502 | 514.901007 | Yes | No | No |
| 28 | 28 | Sacramento Bee | 273.844 | 338.355 | 380.708478 | Yes | No | No |
| 29 | 29 | San Francisco Chronicle | 570.364 | 704.322 | 777.960698 | Yes | No | Yes |
| 30 | 30 | St. Louis Post-Dispatch | 391.286 | 585.681 | 538.047259 | Yes | No | No |
| 31 | 31 | St. Paul Pioneer Press | 201.860 | 267.781 | 284.270451 | Yes | No | No |
| 32 | 32 | Tampa Tribune | 321.626 | 408.343 | 444.722729 | Yes | No | No |
| 33 | 33 | Washington Post | 838.902 | 1165.567 | 1137.725020 | Yes | No | Yes |

Stop Sunday Edition in case of Avg

```
print("Stop Sunday Edition (In case of Avg):")

no_sunday_edition_avg = df[df['Stop_Sunday_Avg'] == 'No']
no_sunday_edition_avg['Newspaper']
```

Stop Sunday Edition (In case of Avg):

```
0          Baltimore Sun
2          Boston Herald
3          Charlotte Observer
6          Cincinnati Enquirer
7          Denver Post
8          Des Moines Register
9          Hartford Courant
11         Kansas City Star
12         Los Angeles Daily News
16         New Orleans Times-Picayune
20         Omaha World Herald
21         Orange County Register
23         Pittsburgh Press
24         Portland Oregonian
25         Providence Journal-Bulletin
26         Rochester Democrat & Chronicle
27         Rocky Mountain News
28         Sacramento Bee
30         St. Louis Post-Dispatch
31         St. Paul Pioneer Press
32         Tampa Tribune
Name: Newspaper, dtype: object
```

Step 6: Desired result

```
print("Stop Sunday Edition (In case of Min):")

no_sunday_edition_min = df[df['Stop_Sunday_Min'] == 'No']
no_sunday_edition_min['Newspaper']
```

Stop Sunday Edition (In case of Min):

Series([], Name: Newspaper, dtype: object)

```
print("Stop Sunday Edition (In case of Max):")

no_sunday_edition_max = df[df['Stop_Sunday_Max'] == 'No']
no_sunday_edition_max['Newspaper']
```

Stop Sunday Edition (In case of Max):

| | |
|----|----------------------------|
| 0 | Baltimore Sun |
| 1 | Boston Globe |
| 2 | Boston Herald |
| 3 | Charlotte Observer |
| 4 | Chicago Sun Times |
| 5 | Chicago Tribune |
| 6 | Cincinnati Enquirer |
| 7 | Denver Post |
| 8 | Des Moines Register |
| 9 | Hartford Courant |
| 10 | Houston Chronicle |
| 11 | Kansas City Star |
| 12 | Los Angeles Daily News |
| 14 | Miami Herald |
| 15 | Minneapolis Star Tribune |
| 16 | New Orleans Times-Picayune |

```
17         New York Daily News
19         Newsday
20         Omaha World Herald
21         Orange County Register
22         Philadelphia Inquirer
23         Pittsburgh Press
24         Portland Oregonian
25         Providence Journal-Bulletin
26         Rochester Democrat & Chronicle
27         Rocky Mountain News
28         Sacramento Bee
29         San Francisco Chronicle
30         St. Louis Post-Dispatch
31         St. Paul Pioneer Press
32         Tampa Tribune
33         Washington Post
Name: Newspaper, dtype: object
```

Linear regression equation for predicting "Sunday" sales based on "Daily" sales:

$$\text{Sunday} = 13.8356 + 1.3397 \cdot \text{Daily}$$

The goal of **linear regression** is to estimate the values of β_0 and β_1 coefficients that best fit the observed data, **minimizing the sum of squared errors** (ϵ^2). This allows us to make predictions or understand the relationship between the dependent and independent variables.