

## TASK 2: SALES PREDICTION

### OUTPUT:

TV advertising spend: [230.1 44.5 17.2 151.5 180.8 8.7 57.5 120.2 8.6 199.8 66.1 214.7

23.8 97.5 204.1 195.4 67.8 281.4 69.2 147.3 218.4 237.4 13.2 228.3

62.3 262.9 142.9 240.1 248.8 70.6 292.9 112.9 97.2 265.6 95.7 290.7

266.9 74.7 43.1 228. 202.5 177. 293.6 206.9 25.1 175.1 89.7 239.9

227.2 66.9 199.8 100.4 216.4 182.6 262.7 198.9 7.3 136.2 210.8 210.7

53.5 261.3 239.3 102.7 131.1 69. 31.5 139.3 237.4 216.8 199.1 109.8

26.8 129.4 213.4 16.9 27.5 120.5 5.4 116. 76.4 239.8 75.3 68.4

213.5 193.2 76.3 110.7 88.3 109.8 134.3 28.6 217.7 250.9 107.4 163.3

197.6 184.9 289.7 135.2 222.4 296.4 280.2 187.9 238.2 137.9 25. 90.4

13.1 255.4 225.8 241.7 175.7 209.6 78.2 75.1 139.2 76.4 125.7 19.4

141.3 18.8 224. 123.1 229.5 87.2 7.8 80.2 220.3 59.6 0.7 265.2

8.4 219.8 36.9 48.3 25.6 273.7 43. 184.9 73.4 193.7 220.5 104.6

96.2 140.3 240.1 243.2 38. 44.7 280.7 121. 197.6 171.3 187.8 4.1

93.9 149.8 11.7 131.7 172.5 85.7 188.4 163.5 117.2 234.5 17.9 206.8

215.4 284.3 50. 164.5 19.6 168.4 222.4 276.9 248.4 170.2 276.7 165.6

156.6 218.5 56.2 287.6 253.8 205. 139.5 191.1 286. 18.7 39.5 75.5

17.2 166.8 149.7 38.2 94.2 177. 283.6 232.1]

Radio advertising spend: [37.8 39.3 45.9 41.3 10.8 48.9 32.8 19.6 2.1 2.6 5.8 24. 35.1 7.6

32.9 47.7 36.6 39.6 20.5 23.9 27.7 5.1 15.9 16.9 12.6 3.5 29.3 16.7

27.1 16. 28.3 17.4 1.5 20. 1.4 4.1 43.8 49.4 26.7 37.7 22.3 33.4

27.7 8.4 25.7 22.5 9.9 41.5 15.8 11.7 3.1 9.6 41.7 46.2 28.8 49.4

28.1 19.2 49.6 29.5 2. 42.7 15.5 29.6 42.8 9.3 24.6 14.5 27.5 43.9

30.6 14.3 33. 5.7 24.6 43.7 1.6 28.5 29.9 7.7 26.7 4.1 20.3 44.5

43. 18.4 27.5 40.6 25.5 47.8 4.9 1.5 33.5 36.5 14. 31.6 3.5 21.

42.3 41.7 4.3 36.3 10.1 17.2 34.3 46.4 11. 0.3 0.4 26.9 8.2 38.  
15.4 20.6 46.8 35. 14.3 0.8 36.9 16. 26.8 21.7 2.4 34.6 32.3 11.8  
38.9 0. 49. 12. 39.6 2.9 27.2 33.5 38.6 47. 39. 28.9 25.9 43.9  
17. 35.4 33.2 5.7 14.8 1.9 7.3 49. 40.3 25.8 13.9 8.4 23.3 39.7  
21.1 11.6 43.5 1.3 36.9 18.4 18.1 35.8 18.1 36.8 14.7 3.4 37.6 5.2  
23.6 10.6 11.6 20.9 20.1 7.1 3.4 48.9 30.2 7.8 2.3 10. 2.6 5.4  
5.7 43. 21.3 45.1 2.1 28.7 13.9 12.1 41.1 10.8 4.1 42. 35.6 3.7  
4.9 9.3 42. 8.6]

Newspaper advertising spend: [ 69.2 45.1 69.3 58.5 58.4 75. 23.5 11.6 1. 21.2 24.2  
4.

65.9 7.2 46. 52.9 114. 55.8 18.3 19.1 53.4 23.5 49.6 26.2  
18.3 19.5 12.6 22.9 22.9 40.8 43.2 38.6 30. 0.3 7.4 8.5  
5. 45.7 35.1 32. 31.6 38.7 1.8 26.4 43.3 31.5 35.7 18.5  
49.9 36.8 34.6 3.6 39.6 58.7 15.9 60. 41.4 16.6 37.7 9.3  
21.4 54.7 27.3 8.4 28.9 0.9 2.2 10.2 11. 27.2 38.7 31.7  
19.3 31.3 13.1 89.4 20.7 14.2 9.4 23.1 22.3 36.9 32.5 35.6  
33.8 65.7 16. 63.2 73.4 51.4 9.3 33. 59. 72.3 10.9 52.9  
5.9 22. 51.2 45.9 49.8 100.9 21.4 17.9 5.3 59. 29.7 23.2  
25.6 5.5 56.5 23.2 2.4 10.7 34.5 52.7 25.6 14.8 79.2 22.3  
46.2 50.4 15.6 12.4 74.2 25.9 50.6 9.2 3.2 43.1 8.7 43.  
2.1 45.1 65.6 8.5 9.3 59.7 20.5 1.7 12.9 75.6 37.9 34.4  
38.9 9. 8.7 44.3 11.9 20.6 37. 48.7 14.2 37.7 9.5 5.7  
50.5 24.3 45.2 34.6 30.7 49.3 25.6 7.4 5.4 84.8 21.6 19.4  
57.6 6.4 18.4 47.4 17. 12.8 13.1 41.8 20.3 35.2 23.7 17.6  
8.3 27.4 29.7 71.8 30. 19.6 26.6 18.2 3.7 23.4 5.8 6.  
31.6 3.6 6. 13.8 8.1 6.4 66.2 8.7]

Sales: [22.1 10.4 9.3 18.5 12.9 7.2 11.8 13.2 4.8 10.6 8.6 17.4 9.2 9.7  
19. 22.4 12.5 24.4 11.3 14.6 18. 12.5 5.6 15.5 9.7 12. 15. 15.9  
18.9 10.5 21.4 11.9 9.6 17.4 9.5 12.8 25.4 14.7 10.1 21.5 16.6 17.1  
20.7 12.9 8.5 14.9 10.6 23.2 14.8 9.7 11.4 10.7 22.6 21.2 20.2 23.7

5.5 13.2 23.8 18.4 8.1 24.2 15.7 14. 18. 9.3 9.5 13.4 18.9 22.3  
 18.3 12.4 8.8 11. 17. 8.7 6.9 14.2 5.3 11. 11.8 12.3 11.3 13.6  
 21.7 15.2 12. 16. 12.9 16.7 11.2 7.3 19.4 22.2 11.5 16.9 11.7 15.5  
 25.4 17.2 11.7 23.8 14.8 14.7 20.7 19.2 7.2 8.7 5.3 19.8 13.4 21.8  
 14.1 15.9 14.6 12.6 12.2 9.4 15.9 6.6 15.5 7. 11.6 15.2 19.7 10.6  
 6.6 8.8 24.7 9.7 1.6 12.7 5.7 19.6 10.8 11.6 9.5 20.8 9.6 20.7  
 10.9 19.2 20.1 10.4 11.4 10.3 13.2 25.4 10.9 10.1 16.1 11.6 16.6 19.  
 15.6 3.2 15.3 10.1 7.3 12.9 14.4 13.3 14.9 18. 11.9 11.9 8. 12.2  
 17.1 15. 8.4 14.5 7.6 11.7 11.5 27. 20.2 11.7 11.8 12.6 10.5 12.2  
 8.7 26.2 17.6 22.6 10.3 17.3 15.9 6.7 10.8 9.9 5.9 19.6 17.3 7.6  
 9.7 12.8 25.5 13.4]

Min sales: 1.6

Sum of sales: 2804.5

Index of min sales: 130

Index of max sales: 175

Variance of sales: 27.085743750000002

Median of sales: 12.9

Mean of sales: 14.0225

Max sales: 27.0

75th percentile of sales: 17.4

Relational Operations:

TV ad spend < 100: [False True True False False True True False True False True False  
 True True False False True False True False False False True False



```
True True True True True True True True True True True  
True True True True True True True True True True True  
True True True True True True True True]
```

Newspaper ad spend > 30: [ True True True True True True False False False False False False False

True	False	True	True	True	True	False	False	True	False	True	False
False	False	False	False	False	True	True	True	False	False	False	False
False	True	True	True	True	True	False	False	True	True	True	False
True	True	True	False	True	True	False	True	True	False	True	False
False	True	False	False	False	False	False	False	False	False	True	True
False	True	False	True	False	False	False	False	False	True	True	True
True	True	False	True	True	True	False	True	True	True	False	True
False	False	True	True	True	True	False	False	False	True	False	False
False	False	True	False	False	False	True	True	False	False	True	False
True	True	False	False	True	False	True	False	False	True	False	True
False	True	True	False	False	True	False	False	False	True	True	True
True	False	False	True	False	False	True	True	False	True	False	False
True	False	True	True	True	True	False	False	False	True	False	False
True	False	False	True	False	False	False	True	False	True	False	False
False	False	False	True	False	False	False	False	False	False	False	False
True	False	False	False	False	False	True	False]				

Sales >= 200: [False False False False False False False False False False False False]

[illegible]

False False False False False False False False False False False False  
False False False False False False False False False False False False  
False False False False False False False False False False False False  
False False False False False False False False False False False False  
False False False False False False False False False False False False  
False False False False False False False False False False False False  
False False False False False False False False False False False False  
False False False False False False False False False False False False

False False False False False False False False]

Sales == 150: [False False False False False False False False False False False False

[illegible]

TV ad spend != 75: [ True True True True True True True True True True True True

True True True True True True True True True True True True  
True True True True True True True True True True True True  
True True True True True True True True True True True True



103.9 159.8 21.7 141.7 182.5 95.7 198.4 173.5 127.2 244.5 27.9 216.8  
 225.4 294.3 60. 174.5 29.6 178.4 232.4 286.9 258.4 180.2 286.7 175.6  
 166.6 228.5 66.2 297.6 263.8 215. 149.5 201.1 296. 28.7 49.5 85.5  
 27.2 176.8 159.7 48.2 104.2 187. 293.6 242.1]

Multiplying radio ad spend by 2: [75.6 78.6 91.8 82.6 21.6 97.8 65.6 39.2 4.2 5.2 11.6 48.  
 70.2 15.2

65.8 95.4 73.2 79.2 41. 47.8 55.4 10.2 31.8 33.8 25.2 7. 58.6 33.4  
 54.2 32. 56.6 34.8 3. 40. 2.8 8.2 87.6 98.8 53.4 75.4 44.6 66.8  
 55.4 16.8 51.4 45. 19.8 83. 31.6 23.4 6.2 19.2 83.4 92.4 57.6 98.8  
 56.2 38.4 99.2 59. 4. 85.4 31. 59.2 85.6 18.6 49.2 29. 55. 87.8  
 61.2 28.6 66. 11.4 49.2 87.4 3.2 57. 59.8 15.4 53.4 8.2 40.6 89.  
 86. 36.8 55. 81.2 51. 95.6 9.8 3. 67. 73. 28. 63.2 7. 42.  
 84.6 83.4 8.6 72.6 20.2 34.4 68.6 92.8 22. 0.6 0.8 53.8 16.4 76.  
 30.8 41.2 93.6 70. 28.6 1.6 73.8 32. 53.6 43.4 4.8 69.2 64.6 23.6  
 77.8 0. 98. 24. 79.2 5.8 54.4 67. 77.2 94. 78. 57.8 51.8 87.8  
 34. 70.8 66.4 11.4 29.6 3.8 14.6 98. 80.6 51.6 27.8 16.8 46.6 79.4  
 42.2 23.2 87. 2.6 73.8 36.8 36.2 71.6 36.2 73.6 29.4 6.8 75.2 10.4  
 47.2 21.2 23.2 41.8 40.2 14.2 6.8 97.8 60.4 15.6 4.6 20. 5.2 10.8  
 11.4 86. 42.6 90.2 4.2 57.4 27.8 24.2 82.2 21.6 8.2 84. 71.2 7.4  
 9.8 18.6 84. 17.2]

Selecting sales values greater than 250: []

Randomly selected sales data:

Unnamed: 0	TV	Radio	Newspaper	Sales
130	131	0.7	39.6	8.7 1.6
199	200	232.1	8.6	8.7 13.4
60	61	53.5	2.0	21.4 8.1
94	95	107.4	14.0	10.9 11.5



5        6   8.7   48.9        75.0   7.2

Vector of 20 evenly spaced values between min and max of Sales:

[ 1.6        2.93684211 4.27368421 5.61052632 6.94736842 8.28421053  
9.62105263 10.95789474 12.29473684 13.63157895 14.96842105 16.30526316  
17.64210526 18.97894737 20.31578947 21.65263158 22.98947368 24.32631579  
25.66315789 27.        ]

Original Sales range: [ 1.6 , 27.0 ]

Vector of 20 evenly spaced normalized values between 0 and 1:

[0.        0.05263158 0.10526316 0.15789474 0.21052632 0.26315789  
0.31578947 0.36842105 0.42105263 0.47368421 0.52631579 0.57894737  
0.63157895 0.68421053 0.73684211 0.78947368 0.84210526 0.89473684  
0.94736842 1.        ]

TV DataFrame:

	TV
0	230.1
1	44.5
2	17.2

```
3  151.5
4  180.8
..  ...
195 38.2
196 94.2
197 177.0
198 283.6
199 232.1
```

[200 rows x 1 columns]

Sales DataFrame:

```
    Sales
0    22.1
1    10.4
2     9.3
3    18.5
4    12.9
..  ...
195   7.6
196   9.7
197  12.8
198  25.5
199  13.4
```

[200 rows x 1 columns]

Joined DataFrame:

TV Sales

0	230.1	22.1
1	44.5	10.4
2	17.2	9.3
3	151.5	18.5
4	180.8	12.9
..	...	...
195	38.2	7.6
196	94.2	9.7
197	177.0	12.8
198	283.6	25.5
199	232.1	13.4

[200 rows x 2 columns]

Pivot table - Average Sales by TV Advertising:

	Sales	
TV		
0.7	1.6	
4.1	3.2	
5.4	5.3	
7.3	5.5	
7.8	6.6	
...	...	
289.7	25.4	
290.7	12.8	
292.9	21.4	
293.6	20.7	

296.4 23.8

[190 rows x 1 columns]

Rows swapped in reverse order:

	Unnamed: 0	TV	Radio	Newspaper	Sales	TV_Range
199	200	232.1	8.6	8.7	13.4	Above 150
198	199	283.6	42.0	66.2	25.5	Above 150
197	198	177.0	9.3	6.4	12.8	Above 150
196	197	94.2	4.9	8.1	9.7	50-100
195	196	38.2	3.7	13.8	7.6	Below 50
..	...	...	...	...	...	...
4	5	180.8	10.8	58.4	12.9	Above 150
3	4	151.5	41.3	58.5	18.5	Above 150
2	3	17.2	45.9	69.3	9.3	Below 50
1	2	44.5	39.3	45.1	10.4	Below 50
0	1	230.1	37.8	69.2	22.1	Above 150

[200 rows x 6 columns]

Columns swapped in reverse order:

	TV_Range	Sales	Newspaper	Radio	TV	Unnamed: 0
0	Above 150	22.1	69.2	37.8	230.1	1
1	Below 50	10.4	45.1	39.3	44.5	2
2	Below 50	9.3	69.3	45.9	17.2	3
3	Above 150	18.5	58.5	41.3	151.5	4
4	Above 150	12.9	58.4	10.8	180.8	5

```

..      ...  ...      ...  ...  ...      ...
195 Below 50  7.6      13.8  3.7  38.2      196
196  50-100  9.7      8.1   4.9  94.2      197
197 Above 150 12.8      6.4   9.3 177.0      198
198 Above 150 25.5      66.2  42.0 283.6      199
199 Above 150 13.4      8.7   8.6 232.1      200

```

[200 rows x 6 columns]

Both rows and columns swapped in reverse order:

```

      TV_Range Sales Newspaper Radio  TV Unnamed: 0
199 Above 150 13.4      8.7   8.6 232.1      200
198 Above 150 25.5      66.2  42.0 283.6      199
197 Above 150 12.8      6.4   9.3 177.0      198
196  50-100  9.7      8.1   4.9  94.2      197
195 Below 50  7.6      13.8  3.7  38.2      196
..      ...  ...      ...  ...  ...      ...
4  Above 150 12.9      58.4  10.8 180.8      5
3  Above 150 18.5      58.5  41.3 151.5      4
2  Below 50  9.3      69.3  45.9  17.2      3
1  Below 50 10.4      45.1  39.3  44.5      2
0  Above 150 22.1      69.2  37.8 230.1      1

```

[200 rows x 6 columns]

Lowercase:

```

0  230.1

```

```
1    44.5
2    17.2
3   151.5
4   180.8
...
195   38.2
196   94.2
197  177.0
198  283.6
199  232.1
```

Name: TV\_str, Length: 200, dtype: object

Length:

```
0    5
1    4
2    4
3    5
4    5
..
195   4
196   4
197   5
198   5
199   5
```

Name: TV\_str, Length: 200, dtype: int64

Starts with '2':

```
0    True
1   False
```

2 False

3 False

4 False

...

195 False

196 False

197 False

198 True

199 True

Name: TV\_str, Length: 200, dtype: bool

Split:

0 [230.1]

1 [44.5]

2 [17.2]

3 [151.5]

4 [180.8]

...

195 [38.2]

196 [94.2]

197 [177.0]

198 [283.6]

199 [232.1]

Name: TV\_str, Length: 200, dtype: object

Extract first digits or letters:

```
0    NaN
1    NaN
2    NaN
3    NaN
4    NaN
```

...

```
195   NaN
196   NaN
197   NaN
198   NaN
199   NaN
```

Name: TV\_str, Length: 200, dtype: object

Find patterns that don't start/end with vowels:

```
0    [230.1]
1    [44.5]
2    [17.2]
3    [151.5]
4    [180.8]
```

...

```
195    [38.2]
196    [94.2]
197   [177.0]
198   [283.6]
199   [232.1]
```

Name: TV\_str, Length: 200, dtype: object

First 3 characters:

```
0    230
```



1 44.

2 17.

3 151

4 180

...

195 38.

196 94.

197 177

198 283

199 232

Name: TV\_str, Length: 200, dtype: object

Last split element:

0 230.1

1 44.5

2 17.2

3 151.5

4 180.8

...

195 38.2

196 94.2

197 177.0

198 283.6

199 232.1

Name: TV\_str, Length: 200, dtype: object

Mean of Sales: 14.0225

Median of Sales: 12.9

Standard deviation of Sales: 5.204396578855228

Variance of Sales: 27.085743750000002

Column: TV

Mean: 147.0425

Median: 149.75

Standard Deviation: 85.63933175679269

Variance: 7334.095143750001

-----

Column: Radio

Mean: 23.264000000000006

Median: 22.9

Standard Deviation: 14.809645640595186

Variance: 219.325604

-----

Column: Newspaper

Mean: 30.553999999999995

Median: 25.75

Standard Deviation: 21.724106057557353

Variance: 471.93678400000005

-----

Column: Sales

Mean: 14.0225

Median: 12.9

Standard Deviation: 5.204396578855228

Variance: 27.085743750000002

-----

Sorted Dataset by Sales:

	TV	Radio	Newspaper	Sales
130	0.7	39.6	8.7	1.6
155	4.1	11.6	5.7	3.2
8	8.6	2.1	1.0	4.8
78	5.4	29.9	9.4	5.3
108	13.1	0.4	25.6	5.3
..	...	...	...	...
36	266.9	43.8	5.0	25.4
147	243.2	49.0	44.3	25.4
198	283.6	42.0	66.2	25.5
183	287.6	43.0	71.8	26.2
175	276.9	48.9	41.8	27.0

[200 rows x 4 columns]

Sorted Dataset by TV, Radio, Newspaper, and Sales:

	TV	Radio	Newspaper	Sales
130	0.7	39.6	8.7	1.6
155	4.1	11.6	5.7	3.2
78	5.4	29.9	9.4	5.3
56	7.3	28.1	41.4	5.5
126	7.8	38.9	50.6	6.6

```

..  ...  ...      ...  ...
98 289.7 42.3    51.2 25.4
35 290.7 4.1     8.5 12.8
30 292.9 28.3   43.2 21.4
42 293.6 27.7    1.8 20.7
101 296.4 36.3   100.9 23.8

```

[200 rows x 4 columns]

Missing Values Before Handling:

Unnamed: 0 0

TV 0

Radio 0

Newspaper 0

Sales 0

TV\_Range 0

TV\_str 0

dtype: int64

Missing Values After Handling:

Unnamed: 0 0

TV 0

Radio 0

Newspaper 0

Sales 0

TV\_Range 0

TV\_str 0

dtype: int64

Filtered Data (Sales > 1000):

Empty DataFrame

Columns: [Unnamed: 0, TV, Radio, Newspaper, Sales, TV\_Range, TV\_str]

Index: []

Numeric columns detected: Index(['Unnamed: 0', 'TV', 'Radio', 'Newspaper', 'Sales'], dtype='object')

Data After Transformation (Scaling):

	Unnamed: 0	TV	Radio	Newspaper	Sales	TV_Range	TV_str
0	-1.723412	0.969852	0.981522	1.778945	1.552053	Above 150	230.1
1	-1.706091	-1.197376	1.082808	0.669579	-0.696046	Below 50	44.5
2	-1.688771	-1.516155	1.528463	1.783549	-0.907406	Below 50	17.2
3	-1.671450	0.052050	1.217855	1.286405	0.860330	Above 150	151.5
4	-1.654129	0.394182	-0.841614	1.281802	-0.215683	Above 150	180.8

