In [1]:
```
import import_ipynb
```

In [11]:
```
import CaseStudy_Solutions
```

Q1 : What data checks will you apply on the given datasets?

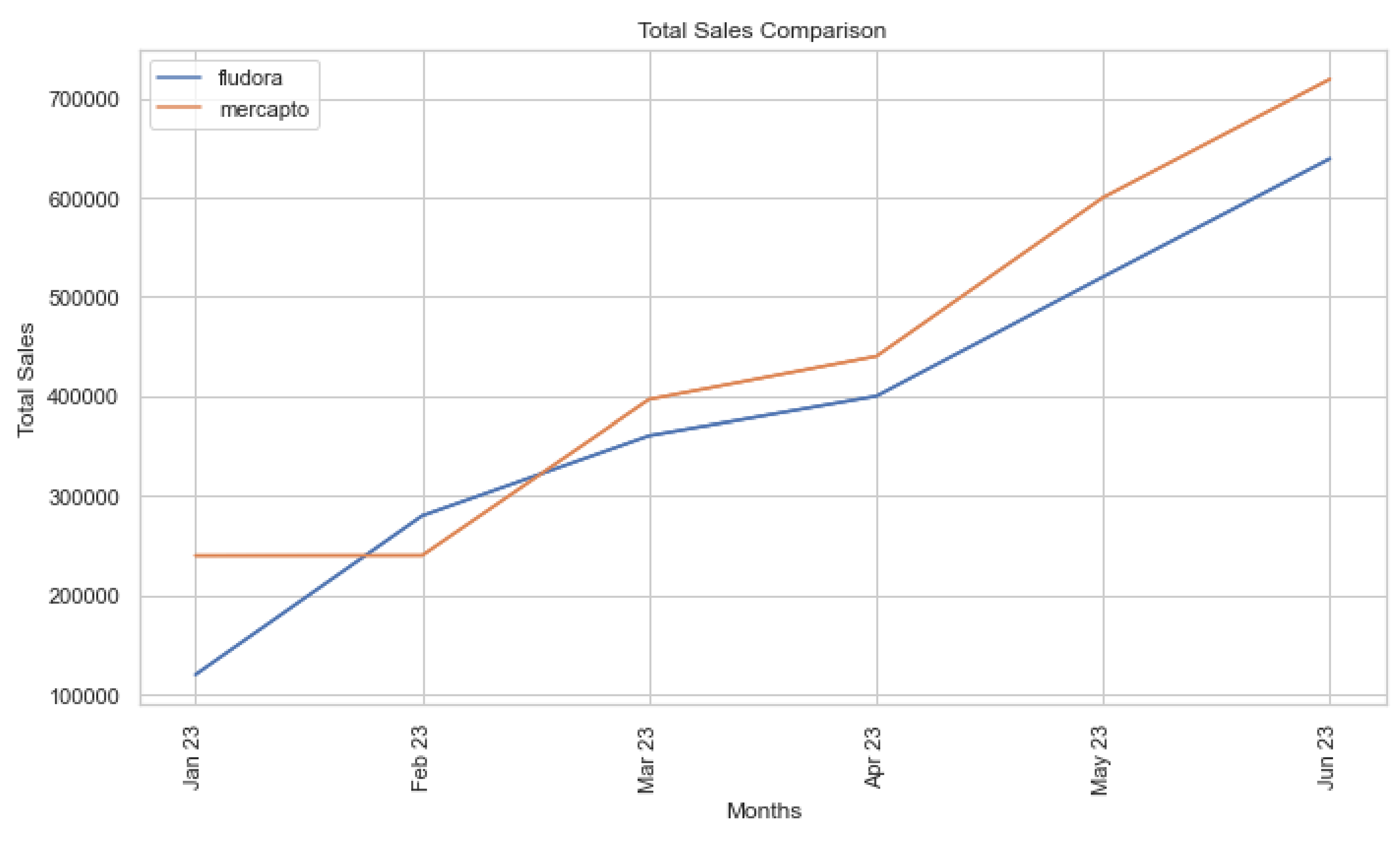
Basically I looked for missing and duplicate value in the data, any removed or ignored them whenever necessary in the questions

Q2 : Plot a graph showing the sales of both the given products (Fludara and Mercapto) over months. Share key insights

In [13]:
```
from CaseStudy_Solutions import dfsales
sns.set(style="whitegrid")

plt.figure(figsize=(10, 6))
sns.lineplot(data=dfsales, dashes=False)
sns.lineplot(data=dfsales, dashes=False)

plt.title('Total Sales Comparison')
plt.xlabel('Months')
plt.ylabel('Total Sales')
plt.legend(labels=dfsales.columns)
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```



Insights that can be drawn : There was sharp increase in growth of Fludara in the first two months but the growth slowed down slighlty after February. This trend continued throughout the year. Meracpto on the other hand had an advantange in sales initially but witnessed almost no growth in sales between Jan to Feb (Probably because Fludara started capturing its market). But After Feb it caught up and maintained a slightly better growth rate than Fludara

Q3 : Who are the top 200 physicians that should be targeted the most? Explain the approach that you considered.

In [3]:
```
from CaseStudy_Solutions import dftarget
dftarget
```

Out[3]:

	Physician ID	Physician Name	Specialty	Jan 23	Feb 23	Mar 23	Apr 23	May 23	Jun 23	Jan 23	...	Jun 23	Total Sales Fludora	Total Sales Mercapto	Total Sales	Gain_Ratio	
0	12757599	RICHARD HYMAS	ORTHOPEDIC SURG, MUSCULOSKELETAL ONCOL	14	16	10	21	30	34	30	...	53	125.0	237.0	362.0	1.896000	68
1	16708862	ASIT JHA	HEMATOLOGY/ONCOLOGY	1	21	30	16	26	30	28	...	54	124.0	227.0	351.0	1.830645	64
2	90321673	Laura Dickerson	HEMATOLOGY	2	17	13	16	29	35	29	...	52	112.0	216.0	328.0	1.928571	63
3	12641680	JAMES SIEBERT	PATHOLOGY, HEMATOLOGY	4	24	10	29	36	33	27	...	51	136.0	233.0	369.0	1.713235	63
4	34897820	JOSHUA DAJAC	SURGICAL ONCOLOGY	0	21	12	21	30	40	22	...	52	124.0	223.0	347.0	1.798387	62
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
195	99792603	Kashif Ali	PATHOLOGY, HEMATOLOGY	2	11	19	18	26	30	16	...	54	106.0	187.0	293.0	1.764151	51
196	35796098	Brian Vikstrom	PEDIATRIC HEMATOLOGY ONCOLOGY	15	18	21	15	26	33	26	...	54	128.0	201.0	329.0	1.570312	51
197	94768562	JESSICA POWERS	ORTHOPEDIC SURG, MUSCULOSKELETAL ONCOL	2	15	13	22	39	37	27	...	51	128.0	201.0	329.0	1.570312	51
198	51889626	MARANGELI FRIGER	RADIOLOGY ONCOLOGY	8	25	31	16	31	33	17	...	47	144.0	210.0	354.0	1.458333	51
199	69785865	MIRA SHAH	HEMATOLOGY/ONCOLOGY	1	22	12	28	36	47	12	...	55	146.0	211.0	357.0	1.445205	51

200 rows × 24 columns

Approach considered:

Firstly, I calculated the sum of total sales for both Fludara and Mercapto for each physician (This will give us the estimate of total volume). In the second part we calculated the ratio of total sales of Mercapto and Fludara (To estimate the scope for increase of Fludaros market share). In the second part we calculated the ratio of total sales of Mercapto and Fludara (To estimate the scope for increase of Fludaros market share). (This would give us the scope of sales we can generate additionally keeping both total sales volume and scope for increase of Fludara market share into consideration.)

Q 4 : How many hospitals don't have any of the top 200 target physicians affiliated to them?

In [4]:
```
from CaseStudy_Solutions import total_hospital
from CaseStudy_Solutions import hospital_with_tardoc
```

In [5]:
```
total_hospital
```

Out[5]:
```
Tennessee Oncology                              27
Mercy Hospital Springfield                      27
Texas Oncology - Mansfield                      26
Cancer And Blood Care Of San Jose Inc           26
Hematology Oncology Associates of Alabama       26
                                                ..
Hematology Oncology                              6
Methodist Hospital                               6
Northwest Oncology And Hematology                5
Windsong Radiology                               4
-                                                3
Name: Hospital Name, Length: 901, dtype: int64
```

In [6]:
```
hospital_with_tardoc
```

Out[6]:
```
Alegent Health Partners                         3
Phoebe Hematology Oncology of Albany            2
Hudson Valley Heme/Onc Associates               2
Regional Hospital of Scranton                   2
Englewood Hospital                              2
                                               ..
Genesis Care - South Florida                    1
De Paul Health Center                           1
Firsthealth Of The Carolinas                    1
Start Center for Cancer Care                    1
Penn Medicine Lancaster General Hospital        1
Name: Hospital Name, Length: 176, dtype: int64
```

First we have the total number of doctors in all the hospitals and in the second list we have total count of target doctors in every hospital. Subtracting them will give us hospitals that dont have any of the top 200 target physicians affiliated to them

901 - 176 = 725

Q5 : List the top 5 hospitals based on the number of Physicians from the following 4 specialties affiliated to them: "Hematology", "Hematology/Oncology", "Oncology Medical" and "Pediatric Hematology Oncology".

In [7]:
```
from CaseStudy_Solutions import hosp_table
hosp_table[:5]
```

Out[7]:

Specialty	HEMATOLOGY	HEMATOLOGY/ONCOLOGY	ONCOLOGY MEDICAL	PEDIATRIC HEMATOLOGY ONCOLOGY	sum
Hospital Name					
OSF Moeller Cancer Center	2	3	5	5	15
Nashville Oncology Associates	2	3	7	2	14
Childrens Hospital And Medical Center Omaha	3	4	3	3	13
Hematology Oncology Associates Of Central New York	3	3	3	4	13
Texas Oncology - Waco	5	3	3	2	13

Q6: Calculate the Workload index for all the territories

In [8]:
```
from CaseStudy_Solutions import sales
sales
```

Out[8]:

	Territory	Region	Total Sales	Workload Index
36	New Jersey N	East	152979.0	1992.811673
47	San Antonio, TX	South	109124.0	1421.525707
26	Los Angeles North, CA	West	109037.0	1420.392384
30	Miami, FL	South	105219.0	1370.656440
6	Charleston, SC	South	104533.0	1361.720123
32	Minneapolis N, MN	West	101271.0	1319.227025
14	Denver, CO	West	99251.0	1292.913089
40	Philadelphia, PA	East	99007.0	1289.734574
20	Indiana	West	97229.0	1266.573100
49	San Francisco N, CA	West	97126.0	1265.231349
15	Detroit, MI	West	97041.0	1264.124080
53	Tampa, FL	South	96735.0	1260.137909
7	Charlotte, NC	East	93253.0	1214.778937
48	San Diego, CA	West	91466.0	1191.500222
24	Las Vegas, NV	West	91141.0	1187.266545
9	Chicago South	West	91000.0	1185.429780
35	New Haven, CT	East	89439.0	1165.095100
25	Long Island, NY	East	85890.0	1118.863338
4	Boston, MA	East	85487.0	1113.613578
31	Milwaukee, WI	West	83776.0	1091.324893
16	Fort Worth, TX	South	83106.0	1082.597003
22	Kansas City, KS	South	81997.0	1068.150392
42	Pittsburgh, PA	East	80845.0	1053.143632
2	Bethesda, MD	East	80661.0	1050.746719
34	Nashville, TN	South	80290.0	1045.913813
11	Cleveland, OH	East	79026.0	1029.448064
8	Chicago North	West	76114.0	991.514311
52	St. Louis, MO	South	75351.0	981.574938
18	Houston, TX	South	74209.0	966.698445
3	Birmingham, AL	South	73852.0	962.047913
39	Orlando, FL	South	73469.0	957.058687
5	Buffalo, NY	East	73341.0	955.391269
27	Los Angeles South, CA	West	72317.0	942.051927
37	New Orleans, LA	South	68271.0	889.345896
46	Roanoke, VA	East	67923.0	884.812604
13	Dallas, TX	South	67488.0	879.145989
41	Phoenix, AZ	West	65930.0	858.850389
28	Madison, WI	West	65811.0	857.300211
17	Harrisburg, PA	East	63283.0	824.368712
21	Jacksonville, FL	South	62414.0	813.048509
45	Richmond, VA	East	60523.0	788.415017
1	Baltimore, MD	East	59574.0	776.052678
0	Atlanta, GA	South	58825.0	766.295679
23	Kentucky	East	58390.0	760.629064
12	Columbus, OH	East	55724.0	725.899880
44	Portland, OR	West	55104.0	717.823325
38	Oklahoma City, OK	South	54900.0	715.165878
51	Seattle S, WA	West	53770.0	700.445706
43	Portland, ME	East	48161.0	627.378941
10	Cincinnati, OH	East	46508.0	605.845804
19	Hudson Valley, NY	East	45949.0	598.563879
50	Seattle N, WA	West	41557.0	541.350608
33	Minneapolis S, MN	West	39060.0	508.822936
29	Manhattan, NY	East	21585.0	281.181338

Q7: Calculate the no. of Territories above and below the balanced workload index range separately. (The territories having a workload index in the range of 700–1,300 (both inclusive) are considered to be balanced)

In [9]:
```
from CaseStudy_Solutions import balanced_table
balanced_table
```

Out[9]:

	Status	Count
0	Balanced	42
1	Unbalanced	12

Q8: Plot a graph depicting the workload index for all the territories in descending order. Which region is performing best based on "Fludara" sales?

In [10]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.set(style="whitegrid")
plt.figure(figsize=(15,6))
sns.barplot(x='Territory', y='Workload Index', data=sales, palette="viridis")  # Create the bar plot

plt.title('Territory vs Workload Index')
plt.xlabel('Territories')
plt.ylabel('Workload Index')
plt.xticks(rotation=90)  # Rotate x-axis labels for better readability
plt.tight_layout()
plt.show()
```