# Pharmaceutical Supply Chain Inventory Optimization: Predicting Energy Consumption

# **Business Overview/Problem**

PharmaCorp faces a significant business challenge related to the optimization of its pharmaceutical supply chain inventory. The primary issues include:

- A. Overproduction and Losses: PharmaCorp often produces medications in quantities that exceed market demand. This overproduction leads to substantial financial losses due to expired shelf-life and the need for disposal of unsold products.
- B. Shelf Life Management: Managing the varying shelf lives of different pharmaceutical products is a complex task. Failure to sell products before their expiration date not only results in financial losses but also poses potential risks to patient safety.
- C. Market Trends: The pharmaceutical industry is highly dynamic, with market trends and demand patterns constantly evolving. PharmaCorp struggles to align its production schedules with these ever-changing market dynamics.
- D. Competitive Pressures: PharmaCorp faces intense competition from other pharmaceutical companies. Efficient inventory management is crucial to maintaining a competitive edge and optimizing costs.

#### **Rationale for the Project**

Rationale for the Project Supply Chain Inventory Optimization is the process of managing inventory levels in a supply chain to minimize costs while meeting customer demand. It is crucial in the pharmaceutical industry because of the perishable nature of drugs and the need to maintain a high level of quality control. The significance of initiating this project lies in several compelling reasons:

- A. Cost Reduction: PharmaCorp incurs significant financial losses due to overproduction and inventory mismanagement. Optimizing inventory based on shelf life and market trends can lead to substantial cost reductions.
- B. Enhanced Profitability: Reducing losses from overproduction and minimizing waste will directly impact PharmaCorp's profitability, allowing resources to be allocated more efficiently.
- C. Competitive Advantage: Efficient inventory management will provide PharmaCorp with a competitive advantage by allowing the

company to respond more effectively to market trends and customer demands.

D. Sustainability: Reducing overproduction and waste aligns with PharmaCorp's commitment to sustainability and responsible business practices.

# Aim of the Project

The project aims to achieve the following objectives:

- Reduce Overproduction: Implement an inventory optimization strategy that significantly reduces overproduction of pharmaceutical products.
- Minimize Losses: Minimize financial losses associated with expired products by aligning inventory with shelf life.
- Improve Forecasting: Enhance demand forecasting accuracy by incorporating market trends and historical sales data.

# **Data Description**

The dataset available from the company contains the following information:

- Product ID: Product unique identifier, for each product.
- Shelf Life Days: Shelf life of the pharmaceutical product in days.
- Sales 2021: Total number of sales of that product in 2021.
- Sales 2022: Total number of sales of that product in 2022.
- Market Trend Factor: An index that measures factors like market trends, consumer preferences, competitor actions, and other external factors.
- Compliance Status: This is an indication of whether or not a drug is compliant with regulation. Can be either of two things 'Compliant' or 'Non-compliant'.

#### **Tools**

#### Programming Language

Python

#### Libraries and Packages

- Pandas
- NumPy

#### Data Visualization

- Matplotlib
- Seaborn

#### Machine Learning:

- Scikit-learn (Sklearn)
- DecisionTreeRegressor
- LabelEncoder
- Metrics (mean\_absolute\_error, mean\_squared\_error, mean\_absolute\_percentage\_error

# 1. Import Packages

# 2. Data Exploration and preprocessing

Manufacturing\_Location

Product\_Category Safety\_Stock\_Days

Storage\_Location dtype: int64

0

0

```
In [2]: 1 df - nd noad csy('Datacet csy')
          1 df hoad()
Out[3]:
             Product_ID Shelf_Life_Days Sales_2021 Sales_2022 Market_Trend_Factor Compliance_Status Supplier_ID Manufacturing_Location
          0
              Product_1
                                   277
                                             602.6
                                                           545
                                                                          0.906303
                                                                                            Compliant
                                                                                                       Supplier_2
                                                                                                                             Location_B
              Product_2
                                   343
                                             359.4
                                                           345
                                                                                                       Supplier_1
                                                                                                                             Location_C
                                                                          0.972500
                                                                                            Compliant
              Product 3
                                                                                                       Supplier_5
                                   291
                                             983.0
                                                          915
                                                                          1.026074
                                                                                        Non-compliant
                                                                                                                             Location_C
                                   298
              Product 4
                                             789.4
                                                           751
                                                                          0.911503
                                                                                            Compliant
                                                                                                       Supplier 8
                                                                                                                             Location C
                                                                                            Compliant Supplier_10
              Product 5
                                   260
                                             326.8
                                                           430
                                                                          1.052617
                                                                                                                             Location A
           1 df icnull() cum()
Out[4]: Product_ID
                                       0
          Shelf_Life_Days
         Sales_2021
          Sales 2022
         Market_Trend_Factor
          Compliance_Status
          Supplier ID
                                       0
```

42.000000

1 df doccniho() Out[5]: Shelf\_Life\_Days Sales\_2021 Sales\_2022 Market\_Trend\_Factor Safety\_Stock\_Days 19381.000000 19381.000000 19381.000000 19381.000000 19381.000000 count 618.423250 618.798256 21.327073 301.737939 0.999851 mean 239.313199 220.590729 0.257548 25.576353 7.090764 std 260.000000 80.000000 234.000000 0.078110 6.000000 min 25% 278.000000 0.823775 429.000000 429.000000 16.000000 50% 302.000000 619.200000 621.000000 1.000239 21.000000 75% 326.000000 811.200000 809.000000 1.175179 27.000000

1.909161

In [6]: 1 df duplicated() all()

max

Out[6]: False

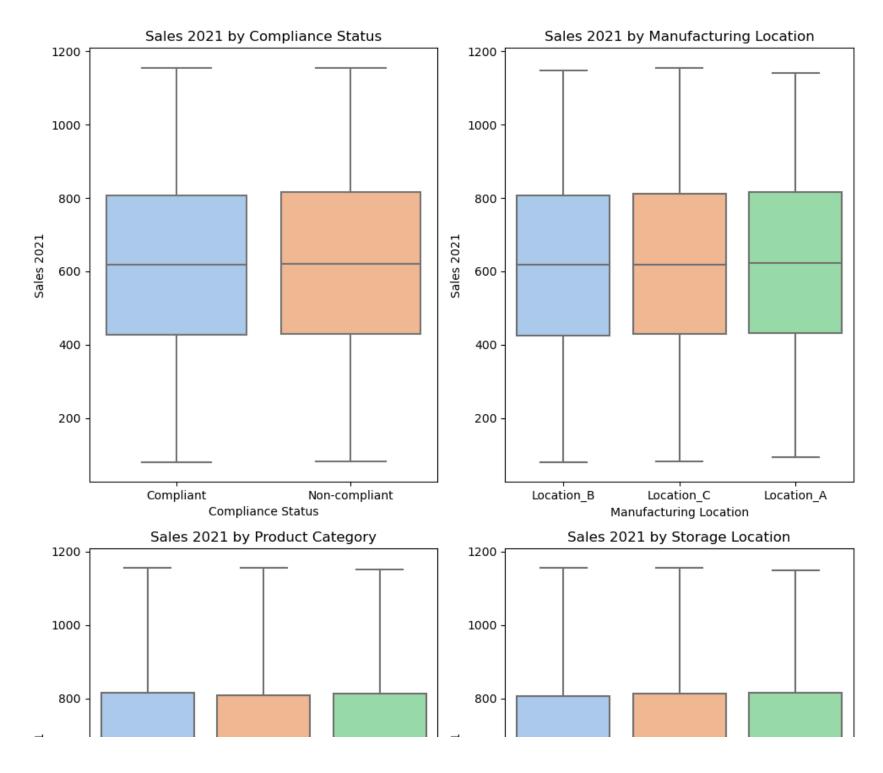
# 2.2 Bivariate Analysis

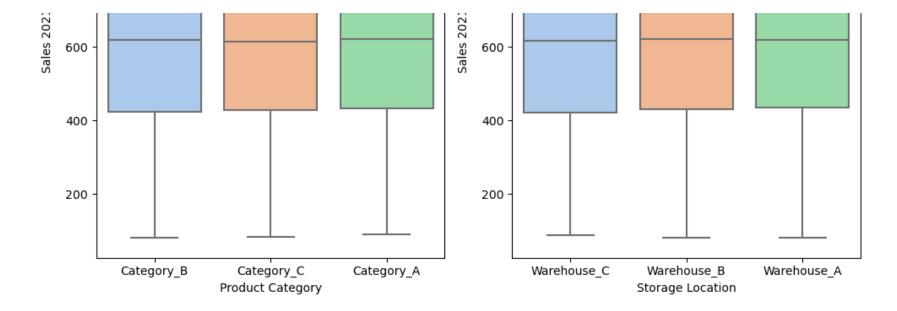
344.000000

1155.400000

999.000000

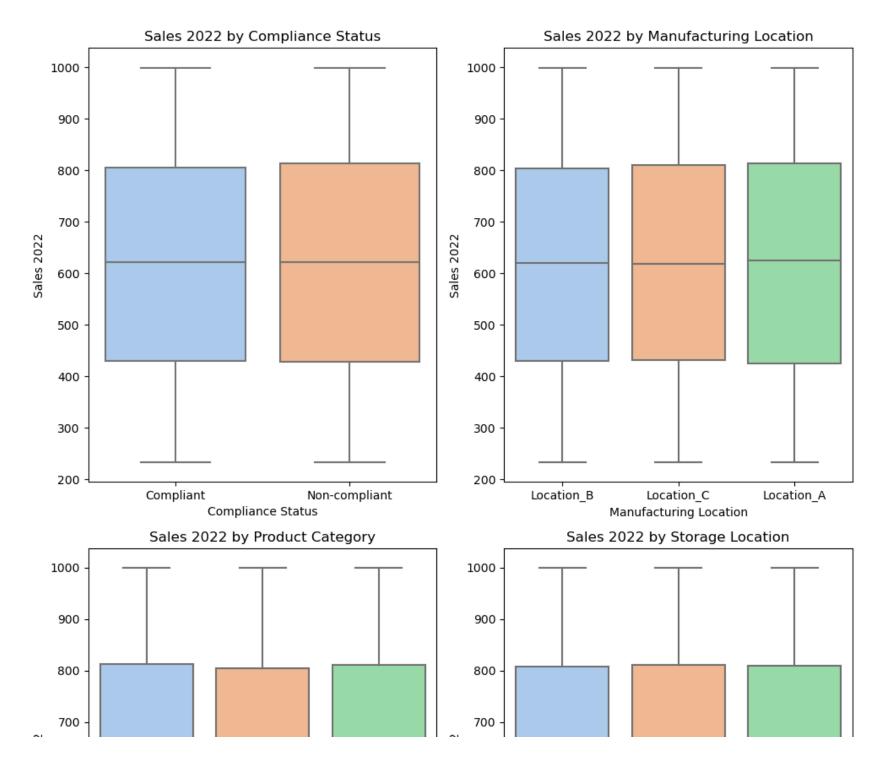
```
In [7]:
         1 import matplotlib.pyplot as plt
          2 import seaborn as sns
         4 fig, ax = plt.subplots(2, 2, figsize=(10, 12))
         6 # Sales 2021 against Compliance Status
         7 sns.boxplot(data=df, x='Compliance_Status', y='Sales_2021', ax=ax[0, 0], palette='pastel')
         8 ax[0, 0].set title('Sales 2021 by Compliance Status')
         9 ax[0, 0].set xlabel('Compliance Status')
        10 ax[0, 0].set ylabel('Sales 2021')
         11
         12 # Sales 2021 against Manufacturing Location
        13 sns.boxplot(data=df, x='Manufacturing_Location', y='Sales_2021', ax=ax[0, 1], palette='pastel')
         14 ax[0, 1].set title('Sales 2021 by Manufacturing Location')
         15 ax[0, 1].set xlabel('Manufacturing Location')
         16 ax[0, 1].set ylabel('Sales 2021')
         17
         18 # Sales 2021 against Product Category
        19 sns.boxplot(data=df, x='Product_Category', y='Sales_2021', ax=ax[1, 0], palette='pastel')
        20 ax[1, 0].set_title('Sales 2021 by Product Category')
         21 ax[1, 0].set xlabel('Product Category')
         22 ax[1, 0].set ylabel('Sales 2021')
         23
         24 # Sales 2021 against Storage Location
        25 sns.boxplot(data=df, x='Storage_Location', y='Sales_2021', ax=ax[1, 1], palette='pastel')
        26 ax[1, 1].set_title('Sales 2021 by Storage Location')
        27 | ax[1, 1].set_xlabel('Storage Location')
         28 ax[1, 1].set ylabel('Sales 2021')
         29
         30 plt.tight_layout()
         31 plt.show()
         22
```

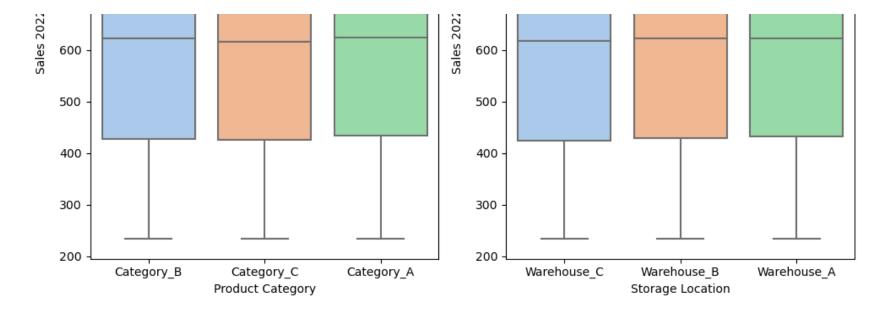




2.2.1 Sales 2022 against Categorical Variables

```
In [8]:
         1 import matplotlib.pyplot as plt
          2 import seaborn as sns
         4 fig, ax = plt.subplots(2, 2, figsize=(10, 12))
         6 # Sales 2022 against Compliance Status
         7 sns.boxplot(data=df, x='Compliance_Status', y='Sales_2022', ax=ax[0, 0], palette='pastel')
         8 ax[0, 0].set title('Sales 2022 by Compliance Status')
         9 ax[0, 0].set xlabel('Compliance Status')
        10 ax[0, 0].set ylabel('Sales 2022')
         11
         12 # Sales 2022 against Manufacturing Location
        13 sns.boxplot(data=df, x='Manufacturing_Location', y='Sales_2022', ax=ax[0, 1], palette='pastel')
         14 ax[0, 1].set title('Sales 2022 by Manufacturing Location')
        15 ax[0, 1].set_xlabel('Manufacturing Location')
         16 ax[0, 1].set ylabel('Sales 2022')
         17
         18 # Sales 2022 against Product Category
        19 sns.boxplot(data=df, x='Product_Category', y='Sales_2022', ax=ax[1, 0], palette='pastel')
        20 ax[1, 0].set_title('Sales 2022 by Product Category')
         21 ax[1, 0].set xlabel('Product Category')
         22 ax[1, 0].set ylabel('Sales 2022')
         23
         24 # Sales 2022 against Storage Location
        25 sns.boxplot(data=df, x='Storage_Location', y='Sales_2022', ax=ax[1, 1], palette='pastel')
        26 ax[1, 1].set_title('Sales 2022 by Storage Location')
        27 | ax[1, 1].set_xlabel('Storage Location')
         28 ax[1, 1].set_ylabel('Sales 2022')
         29
         30 plt.tight_layout()
         31 plt.show()
         22
```

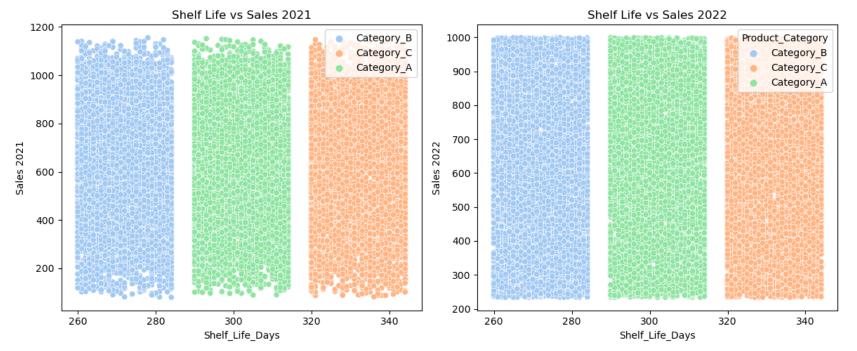




2.2.2 Bivariariate Analysis: sales against Numerical Variables

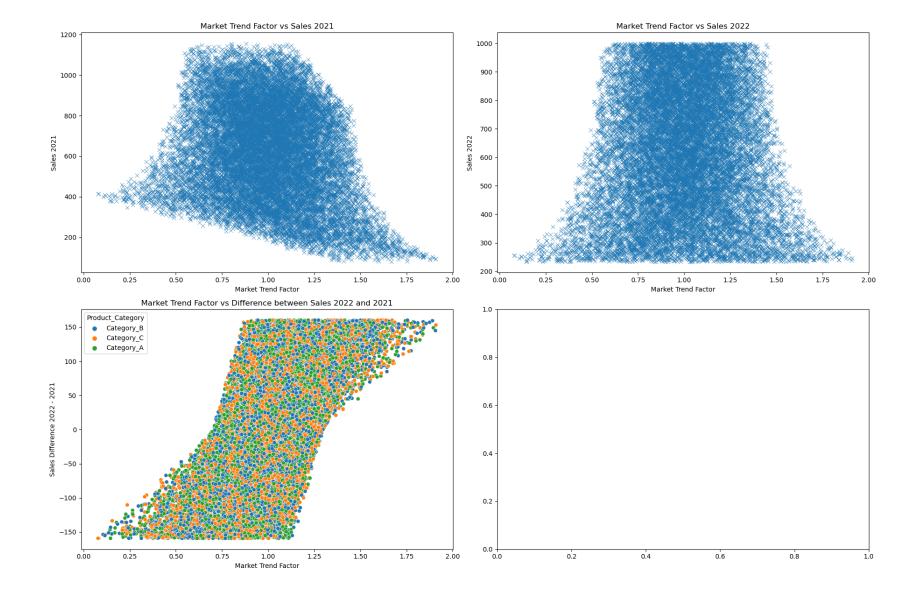
Sales against "Shelf\_life\_Days" and Market\_Trend\_Factor

```
In [9]:
         1 fig, ax = plt.subplots(1, 2, figsize= (12, 5))
          2
            # shelf life Days against Sales 2021
           sns.scatterplot(data=df, x='Shelf_Life_Days', y='Sales_2021', ax=ax[0], hue= 'Product_Category', palette=
           ax[0].set_title('Shelf Life vs Sales 2021')
          6 ax[0].set_xlabel('Shelf_Life_Days')
         7 ax[0].set_ylabel('Sales 2021')
            ax[0].legend(loc = 'upper right')
          9
         10
         11 # shelf life Days against Sales 2022
         12 sns.scatterplot(data=df, x='Shelf_Life_Days', y='Sales_2022', ax=ax[1], hue = 'Product_Category', palette
         13 ax[1].set title('Shelf Life vs Sales 2022')
         14 ax[1].set_xlabel('Shelf_Life_Days')
         15 ax[1].set_ylabel('Sales 2022')
         16 ax[0].legend(loc = 'upper right')
         17
         18 plt.tight_layout()
            plt.show()
         19
         20
```



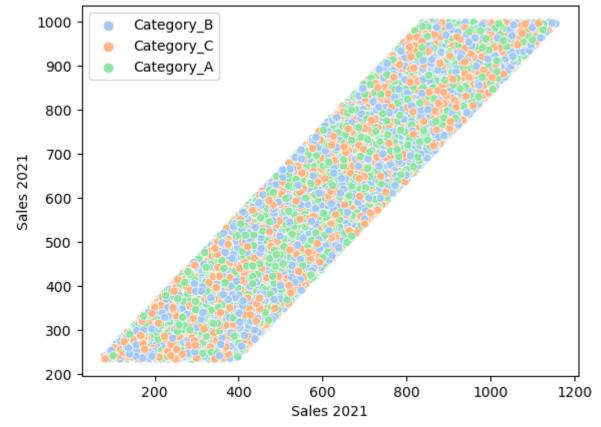
Next Plot for Sales\_2021, Sales\_2022, Market\_Trend\_Factor

```
In [10]:
          1 import matplotlib.pyplot as plt
           2 import seaborn as sns
          4 fig, ax = plt.subplots(2, 2, figsize=(18, 12))
           5
          6 # Market Trend Factor against Sales 2021
          7 | sns.scatterplot(data=df, x='Market_Trend_Factor', y='Sales_2021', ax=ax[0, 0], marker='x')
          8 ax[0, 0].set title('Market Trend Factor vs Sales 2021')
          9 ax[0, 0].set xlabel('Market Trend Factor')
         10 ax[0, 0].set ylabel('Sales 2021')
         11
         12 # Market Trend Factor against Sales 2022
         13 sns.scatterplot(data=df, x='Market_Trend_Factor', y='Sales_2022', ax=ax[0, 1], marker='x')
         14 ax[0, 1].set title('Market Trend Factor vs Sales 2022')
         15 ax[0, 1].set_xlabel('Market Trend Factor')
         16 ax[0, 1].set_ylabel('Sales 2022')
         17
         18 # Market Trend Factor against difference between sales 2022 and 2021
         19 df['Sales_2022-sales_2021'] = df['Sales_2022'] - df['Sales_2021']
         20 sns.scatterplot(data=df, x='Market_Trend_Factor', y='Sales_2022-sales_2021', ax=ax[1, 0], hue='Product_Ca
         21 ax[1, 0].set_title('Market Trend Factor vs Difference between Sales 2022 and 2021')
         22 ax[1, 0].set_xlabel('Market Trend Factor')
         23 ax[1, 0].set_ylabel('Sales Difference 2022 - 2021')
         24
         25 plt.tight_layout()
         26 plt.show()
          27
```



Bivariate Analysis: Sales for 2021 and 2022

# sales 2021 vs Sales 2022



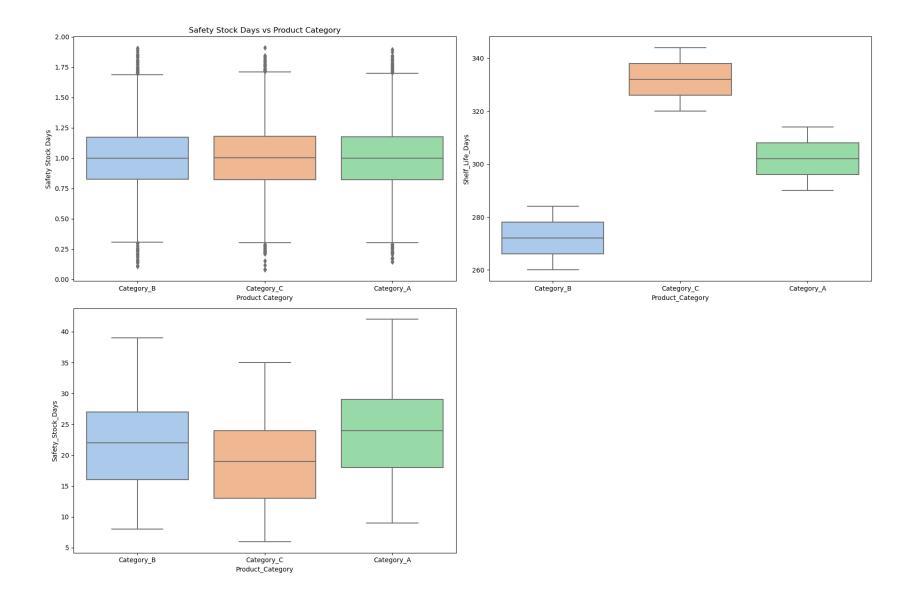
This is a positive correlation

**Bivariate Analysis: Other pairs of variables** 

Product\_Category against

- 'Market\_Trend\_Days',
- 'Safety\_StockDays'

```
1 fig, ax = plt.subplots(2,2, figsize=(18, 12))
In [12]:
          3 # Market Trend Factor and Product Category
          4 sns.boxplot(data=df, x='Product_Category', y='Market_Trend_Factor', ax=ax[0,0], palette= 'pastel')
          5 ax[0,0].set title('Market Trend Factor vs Product Category')
          6 ax[0,0].set_xlabel('Product Category')
          7 ax[0,0].set ylabel('Market Trend Factor')
          8
          9
         10 # Shelf Life days and Product Category
         11 | sns.boxplot(data=df, x='Product_Category', y='Shelf_Life_Days', ax=ax[0,1], palette= 'pastel')
         12 ax[0,0].set title('Shelf Life Days vs Product Category')
         13 ax[0,0].set xlabel('Product Category')
         14 ax[0,0].set_ylabel('Shelf Life Days')
         15
         16
         17 # Safety Stock Days and Product Category
         18 sns.boxplot(data=df, x='Product_Category', y='Safety_Stock_Days', ax=ax[1,0], palette= 'pastel')
         19 ax[0,0].set_title('Safety Stock Days vs Product Category')
         20 ax[0,0].set xlabel('Product Category')
         21 ax[0,0].set ylabel('Safety Stock Days')
         22
         23 ax[1,1].remove()
         24
         25 plt.tight layout()
         26 plt.show()
          27
```



```
In [13]: 1 df info()
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 19381 entries, 0 to 19380
         Data columns (total 12 columns):
              Column
                                      Non-Null Count Dtype
              Product ID
                                      19381 non-null object
              Shelf Life Days
                                      19381 non-null int64
                                      19381 non-null float64
              Sales 2021
              Sales 2022
                                      19381 non-null int64
              Market Trend Factor
                                     19381 non-null float64
              Compliance Status
                                      19381 non-null object
              Supplier ID
                                      19381 non-null object
          7
              Manufacturing_Location 19381 non-null object
              Product Category
                                      19381 non-null
                                                     object
              Safety_Stock_Days
                                      19381 non-null
                                                     int64
          10 Storage Location
                                      19381 non-null
                                                     object
          11 Sales 2022-sales 2021
                                     19381 non-null float64
         dtypes: float64(3), int64(3), object(6)
         memory usage: 1.8+ MB
```

# 3. Demand Forecasting and Inventory Optimization

-Using 'Market\_Trend\_Factor' and Sales\_2022

#### 3.1 Demnd Forecasting and Inventory Optimization; Using Market trend Factor

Market Trend Factor' and Sales 2022

#### **Feature Engineering**

```
In [14]: 1 df_mtf = df.copy()
2
3 df_mtf['Projected_Sales_2023'] = df_mtf['Sales_2022'] * df_mtf['Market_Trend_Factor']
4 df_mtf['Projected_Sales_2023'] = df_mtf['Projected_Sales_2022'] * projected_Sales_2023'] = df_mtf['Projected_Sales_2023'] = df_mtf['Proje
```

#### **Encoding The Categorical Variables**

```
In [15]:
           1 from sklearn.preprocessing import LabelEncoder
           2
           3
             # Columns to encode
             cols = ['Compliance_Status', 'Supplier_ID', 'Manufacturing_Location', 'Product_Category', 'Storage_Locati
           7
             # Label encoding
             label encoder = {}
          10
          11
             for column in cols:
                 le = LabelEncoder()
          12
                 df_mtf[column] = le.fit_transform(df_mtf[column])
          13
                 label encoder[column] = le
          14
In [16]:
          1 df m+f hoad()
```

# Out[16]:

	Product_ID	Shelf_Life_Days	Sales_2021	Sales_2022	Market_Trend_Factor	Compliance_Status	Supplier_ID	Manufacturing_Location
0	Product_1	277	602.6	545	0.906303	0	2	1
1	Product_2	343	359.4	345	0.972500	0	0	2
2	Product_3	291	983.0	915	1.026074	1	5	2
3	Product_4	298	789.4	751	0.911503	0	8	2
4	Product_5	260	326.8	430	1.052617	0	1	0

#### 3.1.2 Prepare X and Y data

```
1 # Splitting into features and target
In [17]:
           2 x = df mtf.drop(columns=['Product ID', 'Projected Sales 2023'])
           2 v - df m+f['Dnoiceted Cales 2022']
In [18]:
          1 | from sklearn.model_selection import train_test_split
           3 # Splitting the data
           4 x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=42)
In [19]:
          1 # Initialize and train the model
           2 model = DecisionTreeRegressor(random_state=42)
           3 model.fit(x_train, y_train)
           5 # Make predictions
           6 y_pred = model.predict(x_test)
           7
           8 mae = mean_absolute_error(y_test, y_pred)
          9 rmse = mean_squared_error(y_test, y_pred)
         10 mape = mean_absolute_percentage_error(y_test, y_pred) * 100
          11
          12
         13 print('MAE:', mae)
         14 print('RMSE:', rmse)
          1E print('MADE.' mana)
         MAE: 6.221055387690978
```

MAE: 6.221055387690978 RMSE: 70.98980614116753 MAPE: 1.1952156870085173

# 3.1.3 Inventory Optimization

Get average daily sales

```
In [20]: 1 df_mtf['Average_Daily_Sales'] = (df_mtf['Sales_2021'] + df_mtf['Sales_2022']) / (2 * 365)
```

Get Safety Stock

```
In [21]: 1 df m+f['Safaty Stock'] - df m+f['Safaty Stock Days'] * df m+f['Ayanaga Daily Salas']
```

#### **Optimal Inventory For 2023**

#### Including Shelf Life In Our Estimation

# 3.2. Demand Forecasting And I nventory Optimization: Linear Projection

#### 3.2.1 Feature Engineering

Generally, the formula is:  $y = y_0 + Slope * (x_1 - x_0)$ 

$$x = YEAR, y = SALES$$

Generally, Slope is :  $Slope = \frac{y_1 - y_0}{x_1 - x_0}$ 

So, for our usecase:

$$Projected\_Sales\_2023 = Sales\_2021 + Slope$$
  
=  $(2023 - 2021)$ 

Type *Markdown* and LaTeX:  $\alpha^2$ 

#### Encode the bcategorical variables

```
In [28]:
           1 cols = ['Compliance_Status',
                     'Supplier_ID',
                      'Manufacturing_Location',
           3
                      'Product_Category',
                      'Storage Location']
           5
             label_encoders = {}
           8
             for column in cols:
                 le = LabelEncoder()
          10
                  df_lp[column] = le.fit_transform(df_lp[column])
          11
                  labal ancodons[column] - la
In [29]:
         1 df ln hoad()
```

#### Out[29]:

	Product_ID	Shelf_Life_Days	Sales_2021	Sales_2022	Market_Trend_Factor	Compliance_Status	Supplier_ID	Manufacturing_Location
0	Product_1	277	602.6	545	0.906303	0	2	1
1	Product_2	343	359.4	345	0.972500	0	0	2
2	Product_3	291	983.0	915	1.026074	1	5	2
3	Product_4	298	789.4	751	0.911503	0	8	2
4	Product 5	260	326.8	430	1.052617	0	1	0

# 3.2.2 Demand Forecasting

#### Build the model.

MAE: 4.609733447979364 RMSE: 5.928638395485055 MAPE: 0.9320903139629664

# 3.2.3. Inventory Optimization

Get Average Daily Sales

```
In [35]:

1 df ln['Avenage Daily Sales'] = (df ln['Sales 2021'] + df ln['Sales 2022']) / (2 * 365)

Get The Safety Stock

In [36]:

1 df ln['Safety Stock'] = df ln['Safety Stock Days'] * df ln['Avenage Daily Sales']

Optimal Inventory for 2023

In [37]:

1 df ln['Optimal Inventory 2023'] = df ln['Decieted Sales 2023'] + df ln['Safety Stock']

Adjust this value to account for shelf life

In [39]:

1 df_lp['Optimal_Inventory_2023'] = df_lp.apply(lambda row: row['Shelf_Life_Days'] * row['Average_Daily_Sales']
```

### 4. Conclusion

Results form Market Trend factor and Sales 2022

In [40]: 1 df mtf[['Doduct ID' 'Dociocted Sales 2022' 'Safety Stock' 'Ontimal Inventory 2022'11 head()
Out[40]:

	Product_ID	Projected_Sales_2023	Safety_Stock	Optimal_Inventory_2023
0	Product_1	493.935278	34.585205	435.459178
1	Product_2	335.512507	23.158356	330.971507
2	Product_3	938.857721	62.400000	756.600000
3	Product_4	684.538694	42.202740	628.820822
4	Product_5	452.625369	21.770959	269.545205

Result from Linear PRojection: Sales 2021 and Sales 2022

```
In [41]: 1 df ln[['Dnoduct TD' 'Dnoiceted Cales 2022' 'Safety Stock' 'Ontimal Inventory 2022'll bead()
Out[41]:
```

	Product_ID	Projected_Sales_2023	Safety_Stock	Optimal_Inventory_2023
0	Product_1	487.4	34.585205	435.459178
1	Product_2	330.6	23.158356	330.971507
2	Product_3	847.0	62.400000	756.600000
3	Product_4	712.6	42.202740	628.820822
4	Product_5	533.2	21.770959	269.545205

```
1 from sklearn.metrics import mean_squared_error
In [46]:
           2
             columns = ['Projected_Sales_2023', 'Optimal_Inventory 2023']
           3
           5
             errors = dict()
           6
             for col in columns:
                  mtf, lp = df_mtf[col], df_lp[col]
           8
           9
                 mae = mean_absolute_error(mtf, lp)
          10
                  rmse = mean_squared_error(mtf, lp,) # Correcting RMSE calculation
          11
          12
                 mape = mean_absolute_percentage_error(mtf, lp) * 100
          13
          14
                  errors[col] = [mae, rmse, mape]
          15
             pd.DataFrame(errors, index=['MAE', 'RMSE', 'MAPE'])
          16
          17
          10
```

# Out[46]:

# MAE 92.516408 21.767899 RMSE 12873.965456 3190.923775 MAPE 16.661423 7.636946

### 5. Recommendations

- 1. Implement an Advanced Predictive Analytics System Action: Adopt a state-of-the-art predictive analytics system that leverages machine learning models to forecast demand accurately. Benefit: This will enable PharmaCorp to anticipate market demand more effectively, reduce overproduction, and minimize the risk of stockouts, leading to more efficient inventory management.
- 2. Develop a Dynamic Shelf Life Management Program Action: Create a dynamic shelf life management program using real-time tracking technologies such as RFID and IoT sensors. Benefit: This program will help PharmaCorp monitor the expiration dates of products more closely, prioritize the sale of near-expiry items, and reduce financial losses due to expired inventory.
- 3. Enhance Supplier Collaboration and Integration Action: Establish stronger collaboration and data integration with suppliers to synchronize supply chain activities. Benefit: Improved communication and data sharing with suppliers can lead to better alignment of production schedules, reduced lead times, and more responsive supply chain operations, ensuring that inventory levels are optimized.
- 4. Adopt Lean Inventory Management Practices Action: Implement lean inventory management practices, such as Just-In-Time (JIT) inventory and continuous improvement (Kaizen) processes. Benefit: These practices will help PharmaCorp reduce excess inventory, lower carrying costs, and enhance operational efficiency by ensuring that inventory levels are closely aligned with actual demand.
- 5. Invest in Employee Training and Development Action: Provide ongoing training and development programs for the supply chain management team on the latest inventory optimization techniques and tools. Benefit: A well-trained team will be better equipped to make data-driven decisions, utilize advanced analytics tools effectively, and continuously improve inventory management processes, leading to enhanced performance and productivity.

In []: