```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib
from matplotlib import pyplot as plt
```

```
excel_file_path = '/content/Projectfinaldata (1) (1) (1).xlsx'
```

```
df = pd.read_excel(excel_file_path)
df
```

	Typeofsales	Patient_ID	Specialisation	Dept	Dateofbill	Quantity	ReturnQuantity	Final_Cost	Final_Sales	RtnMRP	Form
0	Sale	12018098765	Specialisation6	Department1	2022-01-06 00:00:00	1	0	55.406	59.260	0.0	
1	Sale	12018103897	Specialisation7	Department1	7/23/2022	1	0	768.638	950.800	0.0	
2	Sale	12018101123	Specialisation2	Department3	6/23/2022	1	0	774.266	4004.214	0.0	
3	Sale	12018079281	Specialisation40	Department1	3/17/2022	2	0	40.798	81.044	0.0	
4	Sale	12018117928	Specialisation5	Department1	12/21/2022	1	0	40.434	40.504	0.0	
...	...	...	...	...	...	...	...	...	...	...	...
14213	Sale	12018099994	Specialisation39	Department1	6/19/2022	3	0	61.436	145.200	0.0	
14214	Sale	12018047025	Specialisation4	Department1	2/24/2022	2	0	64.448	119.692	0.0	
14215	Sale	12018017139	Specialisation1	Department1	6/27/2022	4	0	74.944	642.040	0.0	
14216	Sale	12018044140	Specialisation20	Department1	7/30/2022	1	0	111.680	181.000	0.0	
14217	Sale	12018116820	Specialisation26	Department1	10/24/2022	3	0	46.182	133.800	0.0	

14218 rows × 14 columns

```
df = pd.read_excel(excel_file_path, sheet_name='Projectfinaldata (1)')
```

```
print(df.head())
```

```
  Typeofsales   Patient_ID    Specialisation         Dept  \
0        Sale  12018098765   Specialisation6  Department1
1        Sale  12018103897   Specialisation7  Department1
2        Sale  12018101123   Specialisation2  Department3
3        Sale  12018079281  Specialisation40  Department1
4        Sale  12018117928   Specialisation5  Department1

            Dateofbill  Quantity  ReturnQuantity  Final_Cost  Final_Sales  \
0  2022-01-06 00:00:00         1               0      55.406       59.260
1            7/23/2022         1               0     768.638      950.800
2            6/23/2022         1               0     774.266     4004.214
3            3/17/2022         2               0      40.798       81.044
4           12/21/2022         1               0      40.434       40.504

   RtnMRP Formulation                          DrugName             SubCat  \
0     0.0       Form1         ZINC ACETATE 20MG/5ML SYP  SYRUP & SUSPENSION
1     0.0       Form1    CEFTAZIDIME 2GM+AVIBACTAM 500MG          INJECTIONS
2     0.0       Form2             EPTIFIBATIDE 0.75MG/ML          INJECTIONS
3     0.0       Form1  WATER FOR INJECTION 10ML SOLUTION          INJECTIONS
```

```
4     0.0       Form1                    LORAZEPAM 1MG  TABLETS & CAPSULES

                                  SubCat1
0                    VITAMINS & MINERALS
1                        ANTI-INFECTIVES
2  CARDIOVASCULAR & HEMATOPOIETIC SYSTEM
3  INTRAVENOUS & OTHER STERILE SOLUTIONS
4                 CENTRAL NERVOUS SYSTEM
```

```
print(df.tail())
```

```
      Typeofsales   Patient_ID    Specialisation         Dept  Dateofbill  \
14213        Sale  12018099994  Specialisation39  Department1   6/19/2022
14214        Sale  12018047025   Specialisation4  Department1   2/24/2022
14215        Sale  12018017139   Specialisation1  Department1   6/27/2022
14216        Sale  12018044140  Specialisation20  Department1   7/30/2022
14217        Sale  12018116820  Specialisation26  Department1  10/24/2022

       Quantity  ReturnQuantity  Final_Cost  Final_Sales  RtnMRP Formulation  \
14213         3               0      61.436      145.200     0.0       Form1
14214         2               0      64.448      119.692     0.0       Form1
14215         4               0      74.944      642.040     0.0       Form1
14216         1               0     111.680      181.000     0.0       Form3
14217         3               0      46.182      133.800     0.0       Form1

                                DrugName                        SubCat  \
14213            SODIUM CHLORIDE IVF 100ML  IV FLUIDS, ELECTROLYTES, TPN
14214  PIPERACILLIN 1GM + TAZOBACTAM 125MG                    INJECTIONS
14215               PARACETAMOL 1GM IV INJ                    INJECTIONS
14216                    MEROPENEM 1GM INJ                    INJECTIONS
14217                             TRAMADOL                    INJECTIONS

                                  SubCat1
14213  INTRAVENOUS & OTHER STERILE SOLUTIONS
14214                        ANTI-INFECTIVES
14215                 CENTRAL NERVOUS SYSTEM
14216                        ANTI-INFECTIVES
14217                 CENTRAL NERVOUS SYSTEM
```

```
#show missing values
missing_values =df.isnull()
print(missing_values)

# Create a heatmap of missing values
plt.figure(figsize=(10, 6))
sns.heatmap(missing_values, cbar=False, cmap='viridis')
plt.title('Missing Values Heatmap')
plt.show()
```

```
       Typeofsales  Patient_ID  Specialisation   Dept  Dateofbill  Quantity  \
0            False       False           False  False       False     False
1            False       False           False  False       False     False
2            False       False           False  False       False     False
3            False       False           False  False       False     False
4            False       False           False  False       False     False
...            ...         ...             ...    ...         ...       ...
14213        False       False           False  False       False     False
14214        False       False           False  False       False     False
14215        False       False           False  False       False     False
14216        False       False           False  False       False     False
14217        False       False           False  False       False     False

       ReturnQuantity  Final_Cost  Final_Sales  RtnMRP  Formulation  DrugName  \
0               False       False        False   False        False     False
1               False       False        False   False        False     False
2               False       False        False   False        False     False
3               False       False        False   False        False     False
4               False       False        False   False        False     False
...               ...         ...          ...     ...          ...       ...
14213           False       False        False   False        False     False
14214           False       False        False   False        False     False
14215           False       False        False   False        False     False
14216           False       False        False   False        False     False
14217           False       False        False   False        False     False

       SubCat  SubCat1
0       False    False
1       False    False
2       False    False
3       False    False
4       False    False
...       ...      ...
14213   False    False
14214   False    False
14215   False    False
14216   False    False
14217   False    False

[14218 rows x 14 columns]
```


```
#remove missing values
df_cleaned = df.dropna()
print(df_cleaned)
```

```
      Typeofsales   Patient_ID    Specialisation         Dept  \
0            Sale  12018098765   Specialisation6  Department1
1            Sale  12018103897   Specialisation7  Department1
2            Sale  12018101123   Specialisation2  Department3
3            Sale  12018079281  Specialisation40  Department1
4            Sale  12018117928   Specialisation5  Department1
...           ...          ...               ...          ...
14213        Sale  12018099994  Specialisation39  Department1
14214        Sale  12018047025   Specialisation4  Department1
14215        Sale  12018017139   Specialisation1  Department1
14216        Sale  12018044140  Specialisation20  Department1
14217        Sale  12018116820  Specialisation26  Department1

                Dateofbill  Quantity  ReturnQuantity  Final_Cost  Final_Sales  \
0      2022-01-06 00:00:00         1               0      55.406       59.260
1                7/23/2022         1               0     768.638      950.800
2                6/23/2022         1               0     774.266     4004.214
3                3/17/2022         2               0      40.798       81.044
4               12/21/2022         1               0      40.434       40.504
...                    ...       ...             ...         ...          ...
14213            6/19/2022         3               0      61.436      145.200
14214            2/24/2022         2               0      64.448      119.692
14215            6/27/2022         4               0      74.944      642.040
14216            7/30/2022         1               0     111.680      181.000
14217           10/24/2022         3               0      46.182      133.800

       RtnMRP Formulation                          DrugName  \
0         0.0       Form1           ZINC ACETATE 20MG/5ML SYP
1         0.0       Form1     CEFTAZIDIME 2GM+AVIBACTAM 500MG
2         0.0       Form2              EPTIFIBATIDE 0.75MG/ML
3         0.0       Form1   WATER FOR INJECTION 10ML SOLUTION
4         0.0       Form1                       LORAZEPAM 1MG
...       ...         ...                                 ...
14213     0.0       Form1           SODIUM CHLORIDE IVF 100ML
14214     0.0       Form1  PIPERACILLIN 1GM + TAZOBACTAM 125MG
```

```
14215     0.0       Form1              PARACETAMOL 1GM IV INJ
14216     0.0       Form3                   MEROPENEM 1GM INJ
14217     0.0       Form1                          TRAMADOL

                            SubCat                            SubCat1
0                SYRUP & SUSPENSION                VITAMINS & MINERALS
1                        INJECTIONS                    ANTI-INFECTIVES
2                        INJECTIONS  CARDIOVASCULAR & HEMATOPOIETIC SYSTEM
3                        INJECTIONS  INTRAVENOUS & OTHER STERILE SOLUTIONS
4                TABLETS & CAPSULES              CENTRAL NERVOUS SYSTEM
...                             ...                               ...
14213  IV FLUIDS, ELECTROLYTES, TPN  INTRAVENOUS & OTHER STERILE SOLUTIONS
14214                    INJECTIONS                    ANTI-INFECTIVES
14215                    INJECTIONS              CENTRAL NERVOUS SYSTEM
14216                    INJECTIONS                    ANTI-INFECTIVES
14217                    INJECTIONS              CENTRAL NERVOUS SYSTEM

[12037 rows x 14 columns]
```

```
# Set the figure size
plt.figure(figsize=(10, 6))

# Plot the bars
plt.bar(df_cleaned['Patient_ID'], df_cleaned['Final_Sales'], label='Final_Sales', color='blue',width=100)
plt.bar(df_cleaned['Patient_ID'],df_cleaned['ReturnQuantity'], label='ReturnQuantity', color='red',width=100)

plt.xlabel('Patient_ID')
plt.ylabel('Values')
plt.title('Final_Sales vs. ReturnQuantity by Product')
plt.legend()

# Show the plot
plt.xticks(rotation=45)  # Rotate the x-axis labels for better readability
plt.tight_layout()  # Ensure the labels fit within the figure
plt.show()
```