**Business Problem:**

To do analysis on Medical inventory optimization based on drugs and to Inventory deals with warehouses for the storage of drugs according to the condition of the patients and the number of drugs being sold. By forecasting we can optimize the inventory for future and past .

**Business Objectives:**

1.Improving stock management and reducing waste: Accurately predicting demand for medicine can help ensure that there is always enough inventory on hand, while reducing the amount of waste due to overstocking or expiration of unused products.

2.Increasing efficiency and reducing costs: An effective forecasting system can help streamline operations and reduce the costs associated with managing medicine inventory.

3.Minimise the Zero bounce

4.Enhancing customer satisfaction: By ensuring that the right medicines are always in stock, a forecasting system can help improve customer satisfaction and increase loyalty.

5.To optimise inventory and minimise cash drain, ultimately saving on inventory cost

**Business Constraints:**

1.Data availability and quality: Forecasting requires accurate and up-to-date data, accuracy depends on the quality of data

2.Uncertainty in future events: Forecasting is inherently uncertain, and the accuracy of predictions can be impacted by unexpected events such as natural disasters, pandemics, and economic downturns.

**Data Pre-Processing:**

* Shape of the data (14218 X 14)
* Missing values before imputation: 5681
* Missing values after imputation: 0
* Number of Duplicate records is 15 and we drop them.

**Exploratory Data Analysis :**

**First moment business decision**

**#1st Business Moment : Measures of Central Tendency -> Mean, Median, Mode**

**MEAN**

PATIENT\_ID 1.201809e+10

QUANTITY 2.231748e+00

RETURN\_QUANTITY 2.919539e-01

FINAL\_COST 1.248240e+02

FINAL\_SALES 2.340383e+02

RETURN\_MRP 2.912675e+01

dtype: float64

**MEDIAN**

PATIENT\_ID 1.201809e+10

QUANTITY 1.000000e+00

RETURN\_QUANTITY 0.000000e+00

FINAL\_COST 5.365000e+01

FINAL\_SALES 8.642400e+01

RETURN\_MRP 0.000000e+00

dtype: float64

**Second moment bussiness decision**

**Measures of Dispersion -> Variance, Standard Deviation, Range**

**VARIANCE**

PATIENT\_ID 7.897389e+08

QUANTITY 2.633786e+01

RETURN\_QUANTITY 2.700506e+00

FINAL\_COST 2.160230e+05

FINAL\_SALES 4.505921e+05

RETURN\_MRP 3.321956e+04

dtype: float64

**Standard deviation**

PATIENT\_ID 28102.293889

QUANTITY 5.132043

RETURN\_QUANTITY 1.643322

FINAL\_COST 464.782794

FINAL\_SALES 671.261572

RETURN\_MRP 182.262335

dtype: float64

**# Third moment bussiness decision**

SKEWNESS

PATIENT\_ID -1.366038

QUANTITY 11.341034

RETURN\_QUANTITY 17.172365

FINAL\_COST 34.508215

FINAL\_SALES 21.006722

RETURN\_MRP 15.797853

dtype: float64

**Fourth moment bussiness decision**

**KURTOSIS**

15

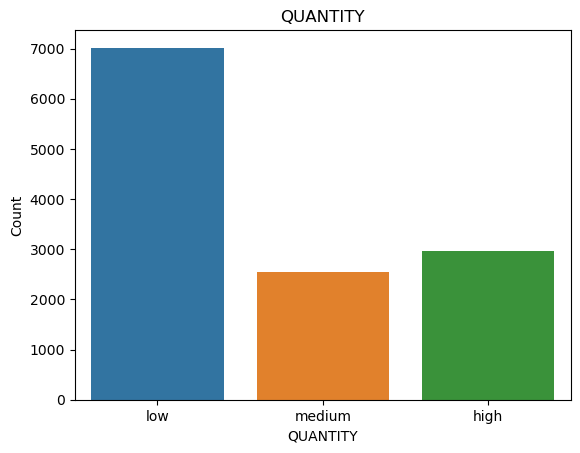
**OUTLIERS**

* Outliers found in quantity, return quantity, final cost, final sales and return MRP columns.
* Created bins according to the percentile.

LOW - 0 to 25th percentile

MEDIUM -25th to 75th percentile

HIGH -75th to 100 percentile



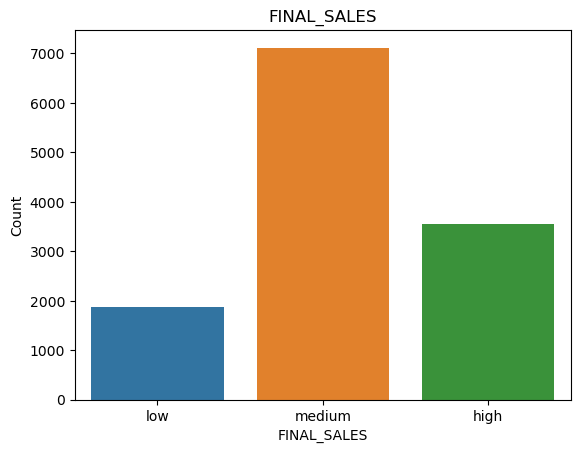
This plot refers quantity vs count

1.The first bin includes values between the minimum value and the 25th percentile

2.The second bin includes values between the 25th and 75th percentiles

3.The third bin includes values between the 75th percentile and the maximum value.

Hence, low is maximum.



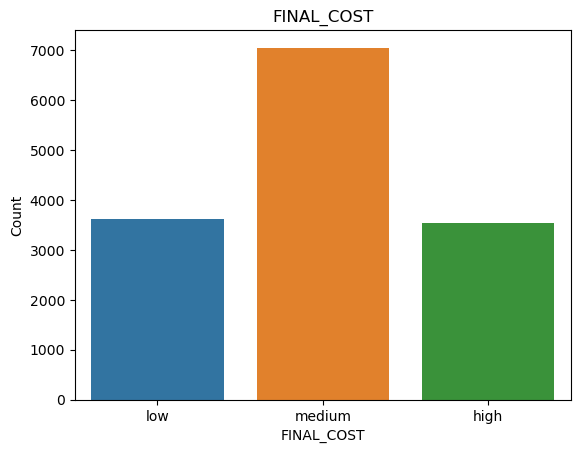
This plot refers FINAL\_SALES vs count

1.The first bin includes values between the minimum value and the 25th percentile

2.The second bin includes values between the 25th and 75th percentiles

3.The third bin includes values between the 75th percentile and the maximum value.

Hence, medium is maximum.



This plot refers FINAL\_COST vs count

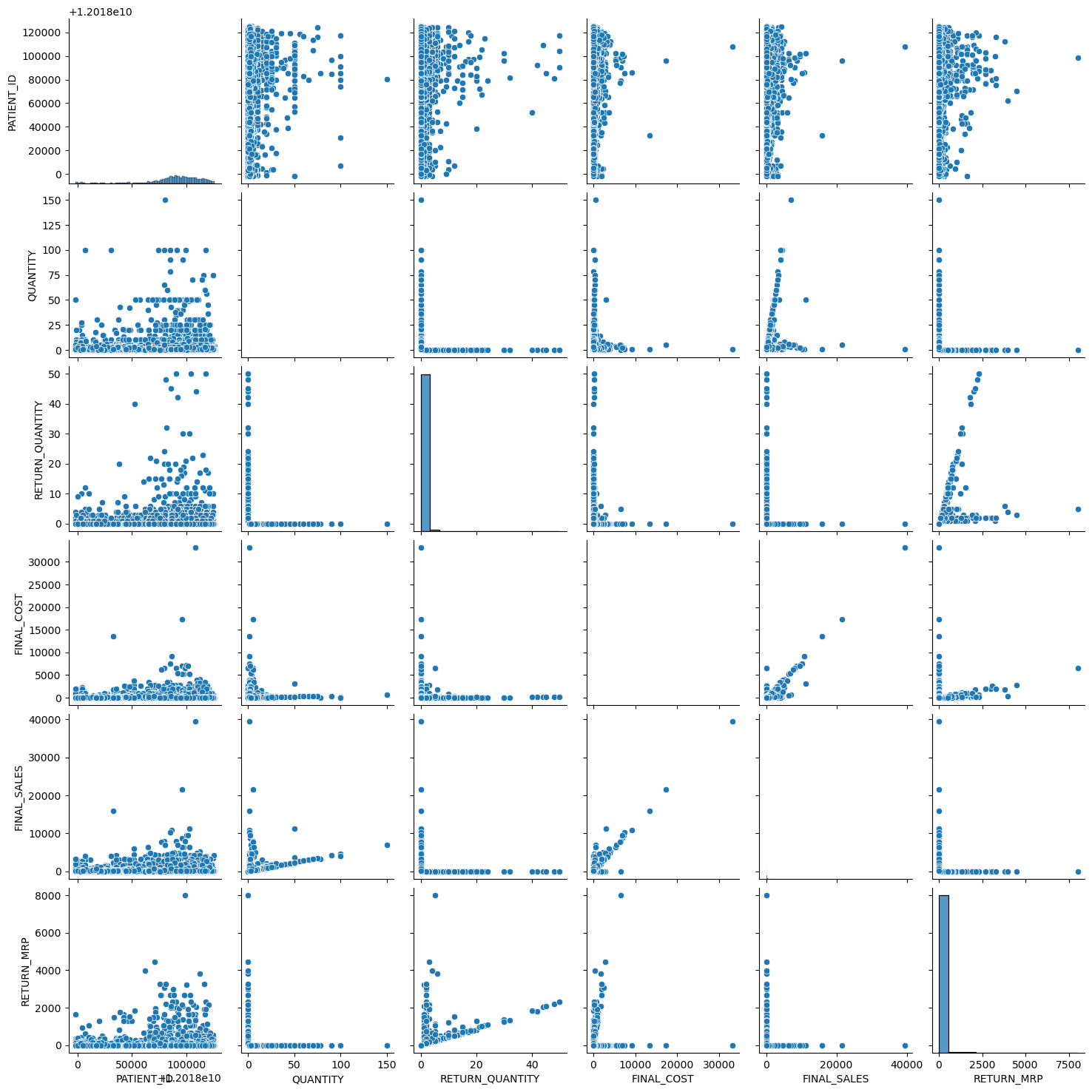
1.The first bin includes values between the minimum value and the 25th percentile

2.The second bin includes values between the 25th and 75th percentiles

3.The third bin includes values between the 75th percentile and the maximum value.

Hence, medium is maximum.

SEABORN PAIRPLOT



1.By importing the seaborn library we use to see the grid of scatter plots

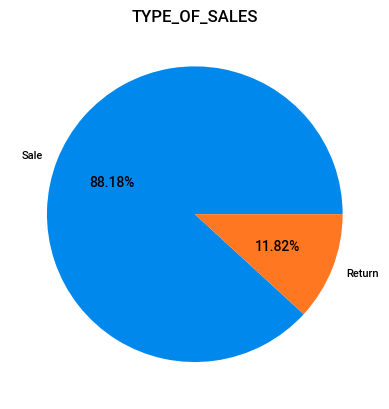
2.Each plot in the grid shows the relationship between two variables.

3. The variables are plotted on the x and y axis, and each data point in the

plot represents the values of the two variables for a single observation in the DataFrame.

**Exploratory Data Analysis (EDA):**

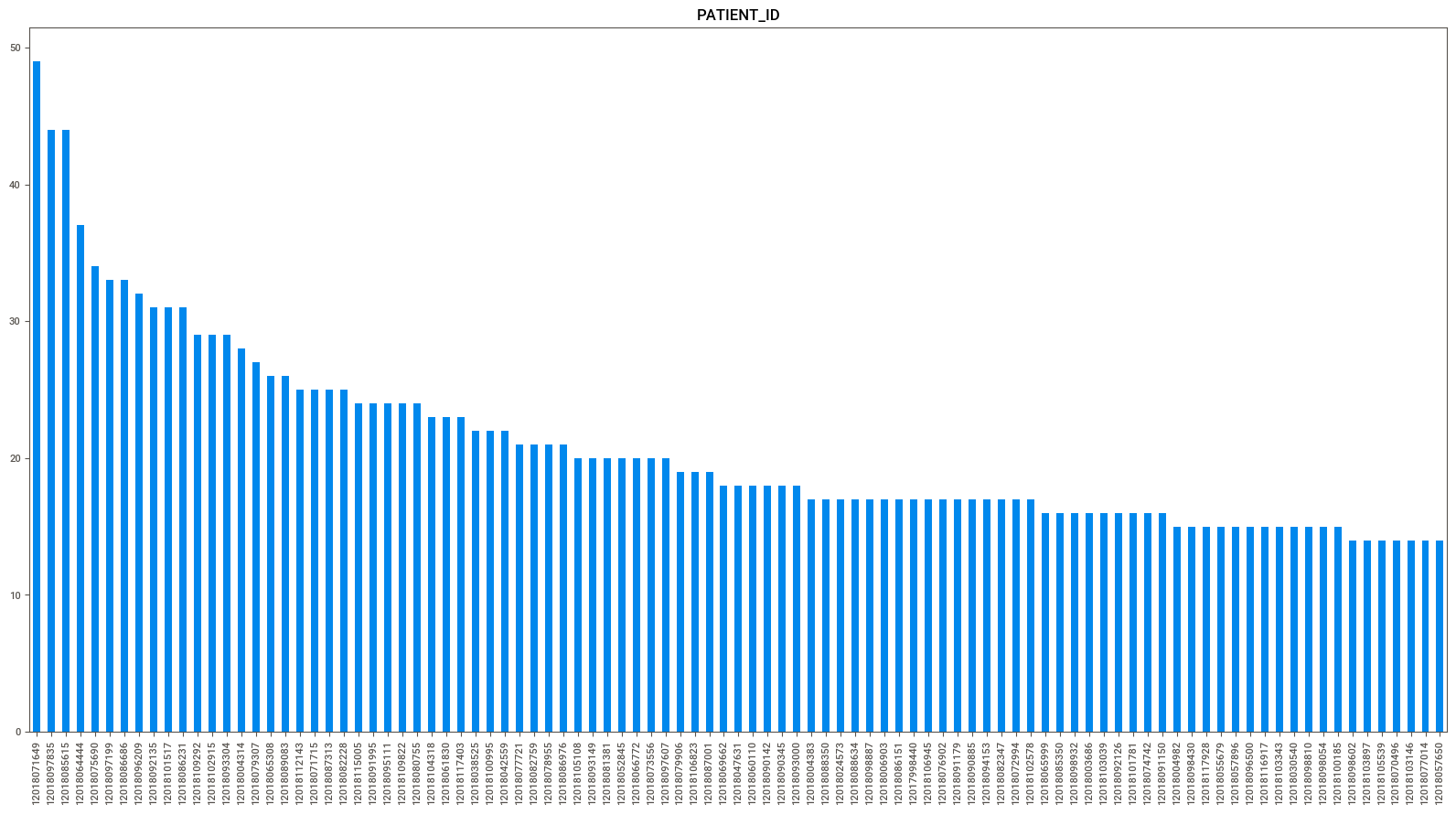
**Type of sales column:**



1.In this pie chart it represents two unique values that is sales and returns

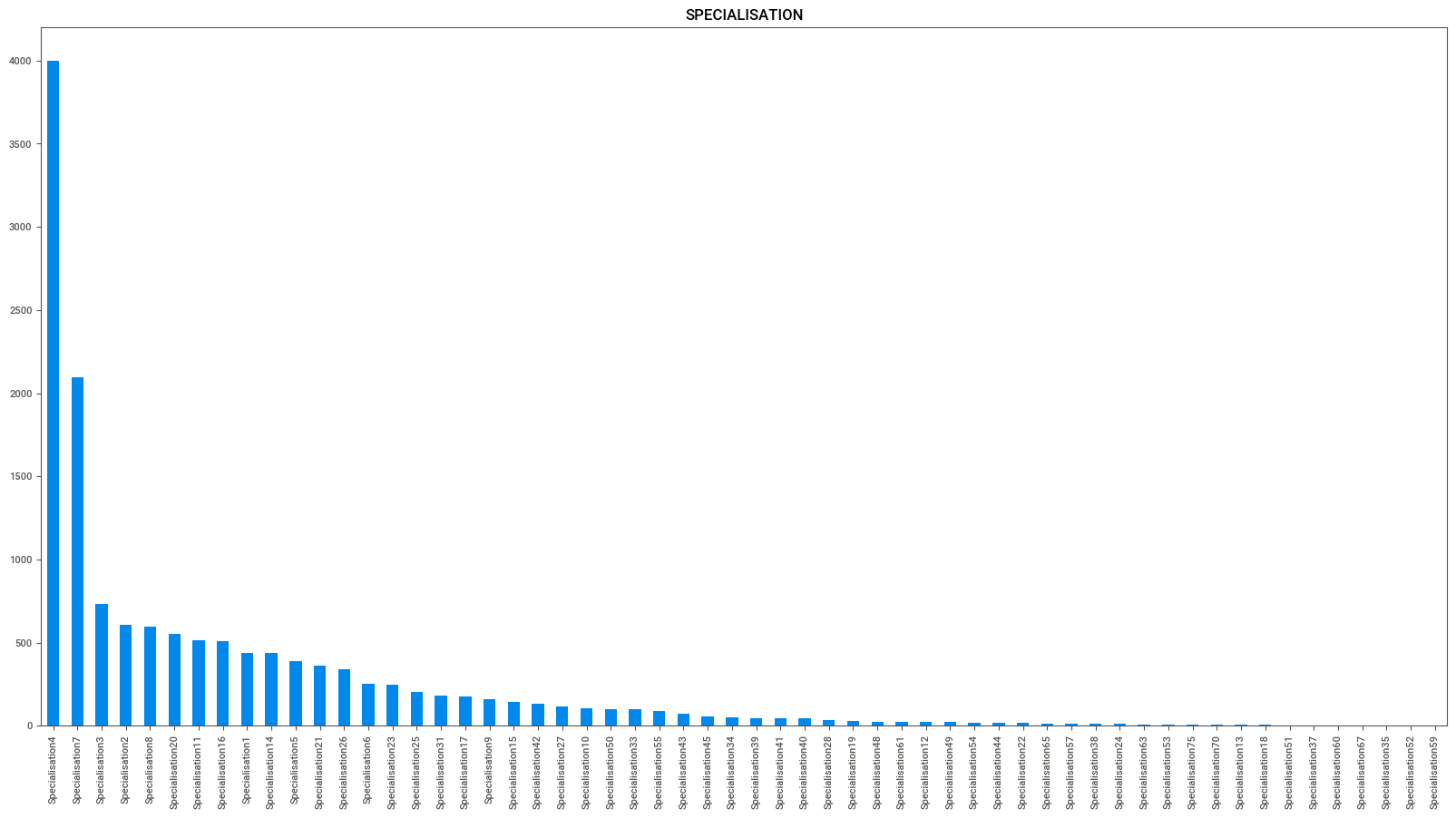
2.Sales happened 88.19% and return is 11.81%.

**Patient ID column:**



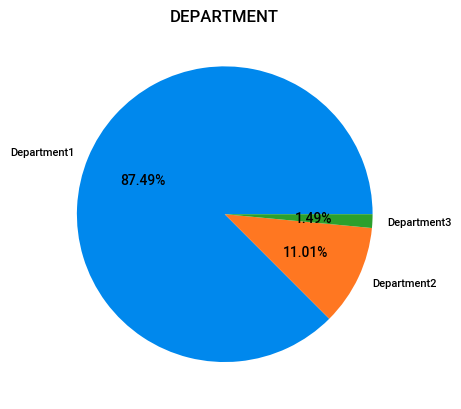
In this bar chart we see the most number of patients visited to the hospital.

**Specialisation column:**

****

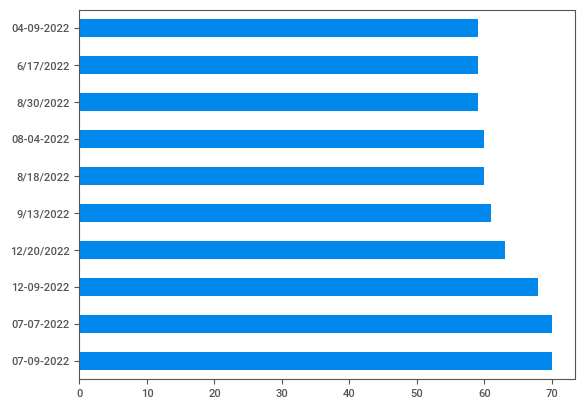
* In this chart it represents 58 specialisation in the hospitals
* Patient’s undergoing for treatments in specialisation 4 is high, followed by specialisation 7.
* Patient’s least undergoing specialisation 63, 53,75,70,13,18,51,37,60,67,35,52 and 59.

**Department Column:**

****

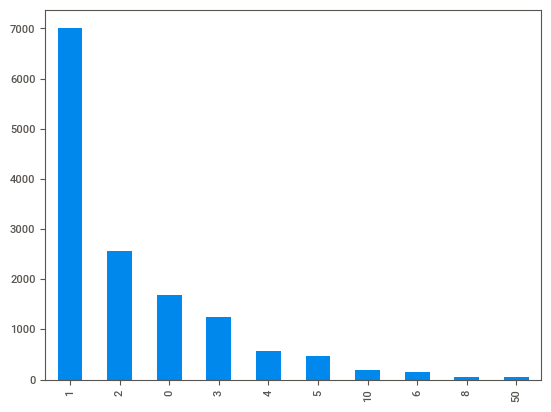
This chart represents there are 3 departments in the hospital. Department 1 is contributing the most, followed by department2 and then by department3.

**Date of billing Column:**

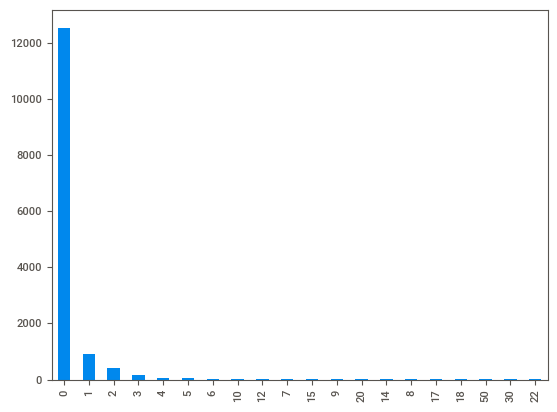
****

* This is 1 year data which includes sales and return of drugs.

quantity column

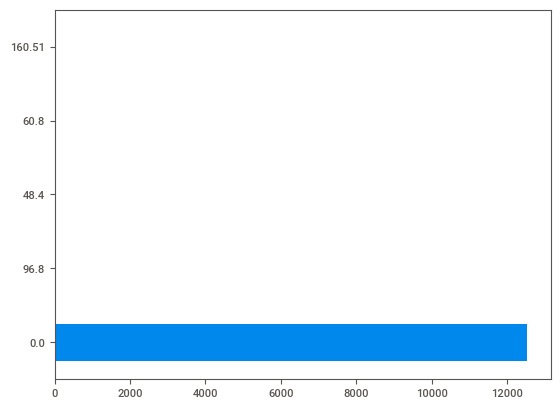


The unique values in quantity is 45.  
**return quantity**

****

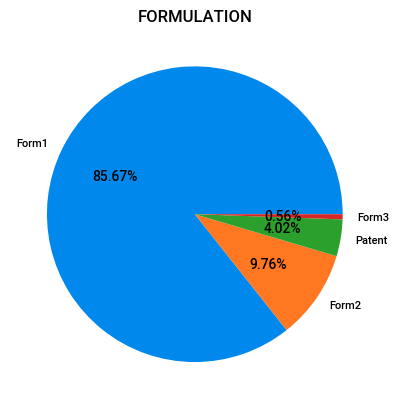
**The unique values in Return quantity is 33.**

**Return\_MRP**

****

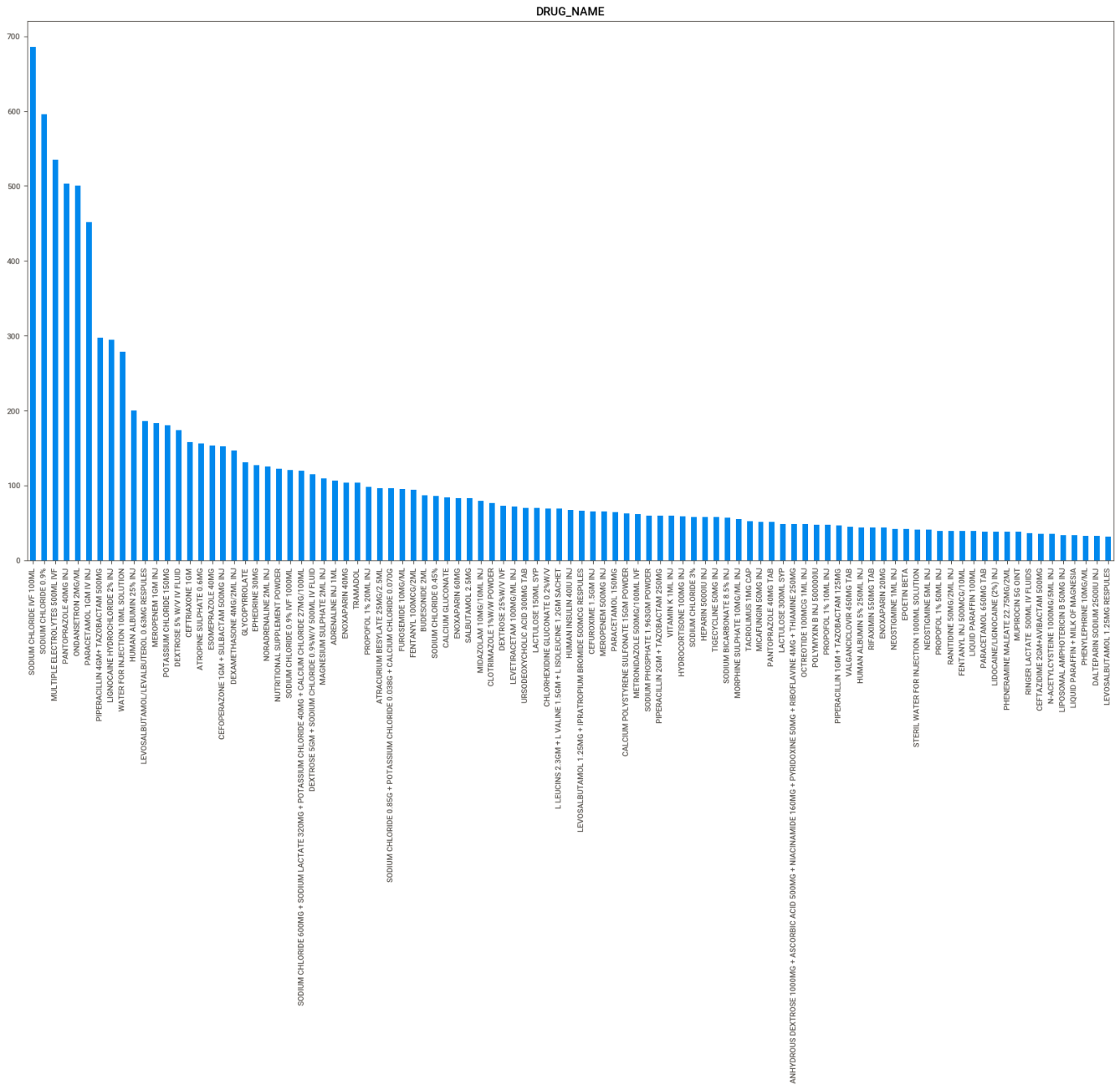
**The unique values in RETURN\_MRP is 689.**

**Formulation Column:**

****

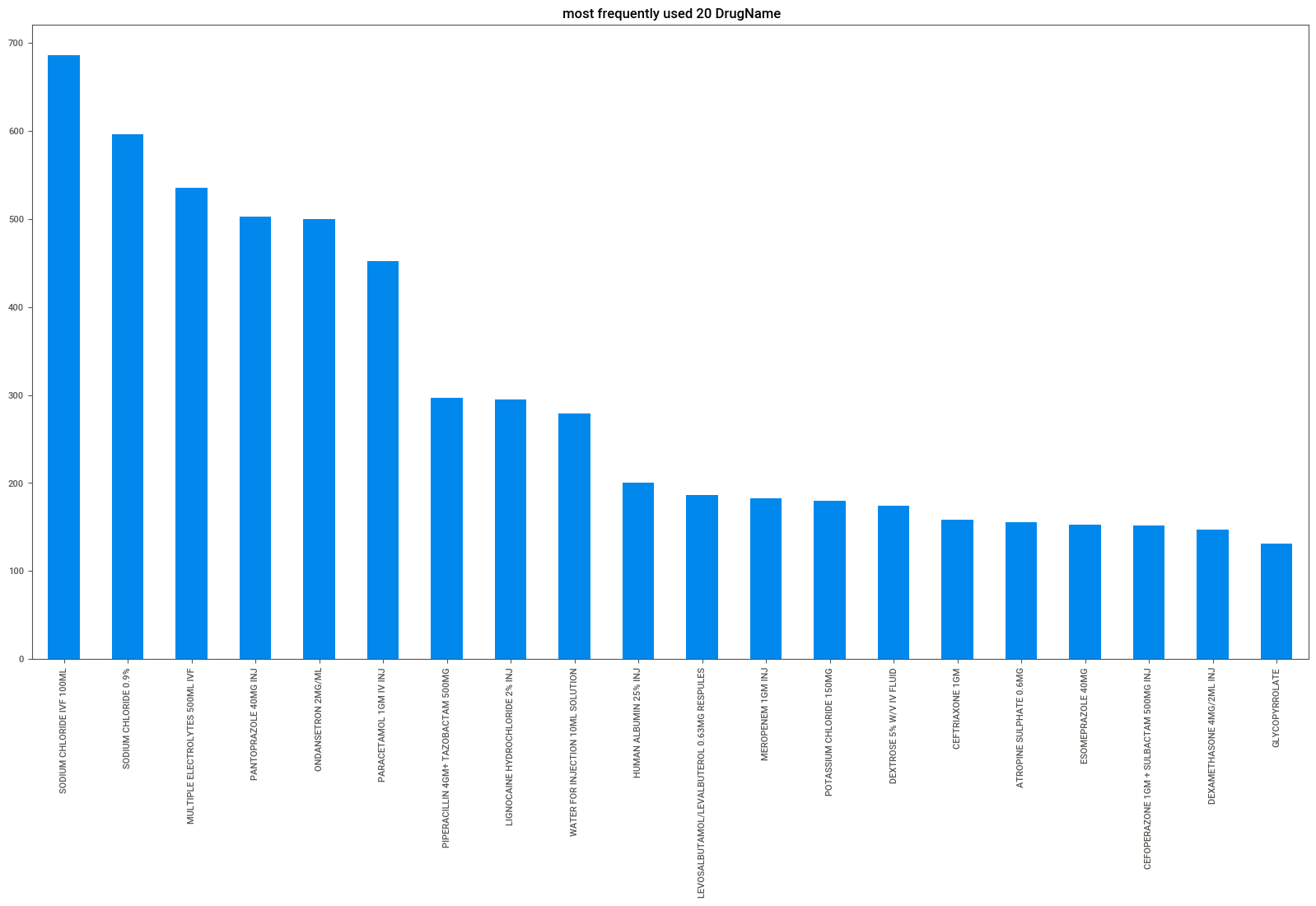
In this pie chart it represents that Form1 contributes the most, followed by form2, Patent and form3

**DrugName column:**

****

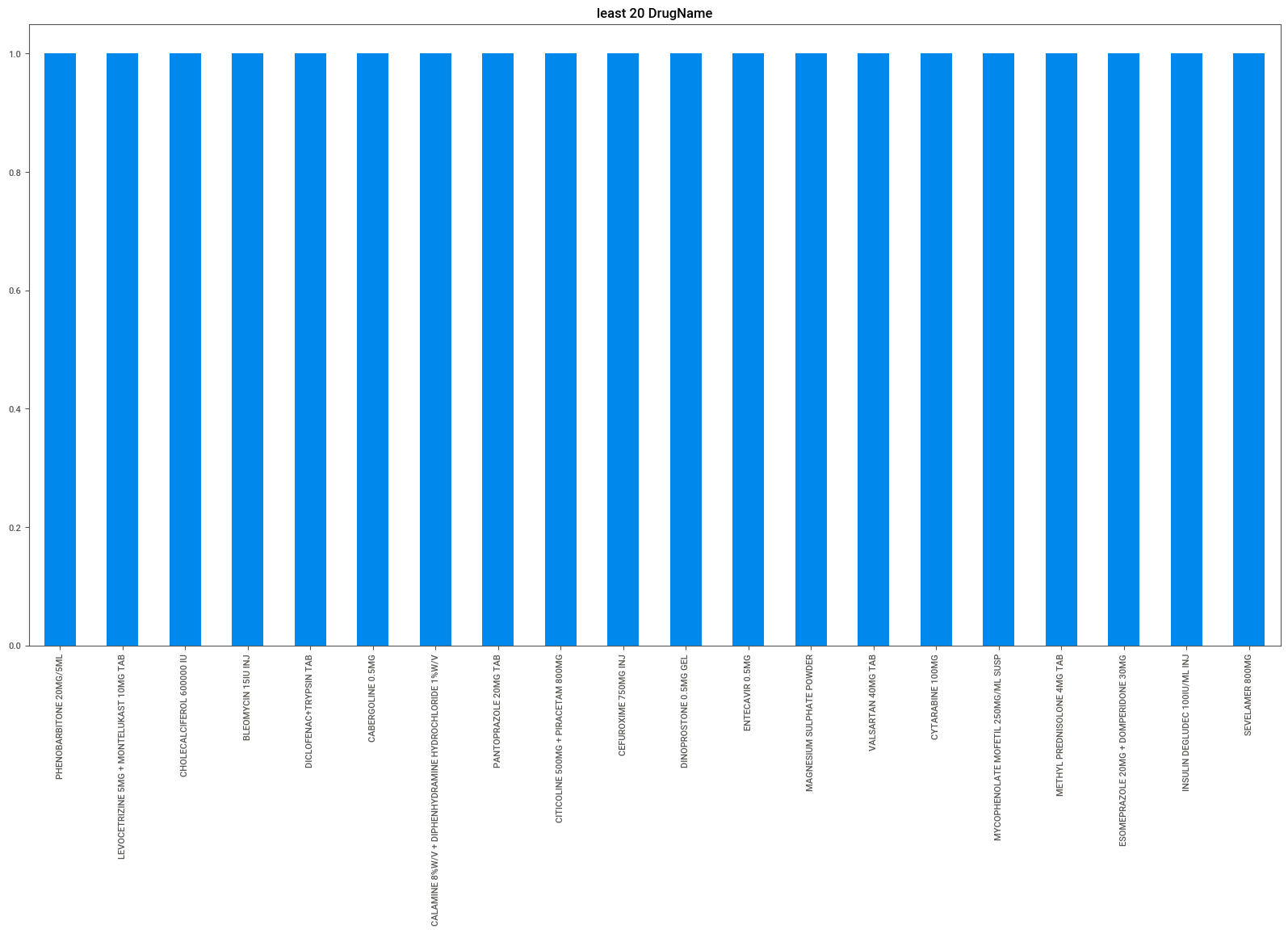
* Total 752 unique drugs

**Top most frequently used 20 drug based on their count.**



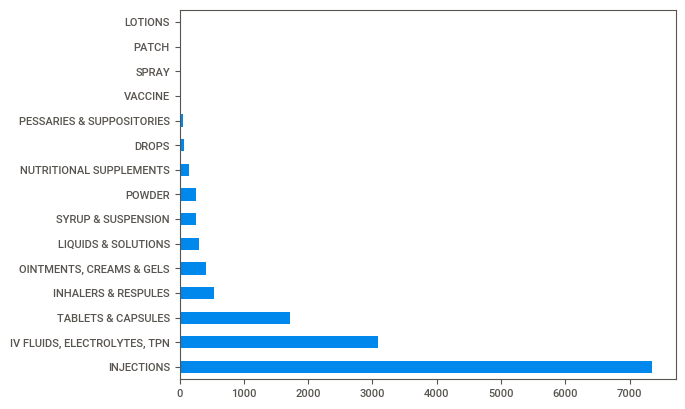
This bar chart represents top most frequently used 20 drugs

**Least used 20 drugs based on their count.**



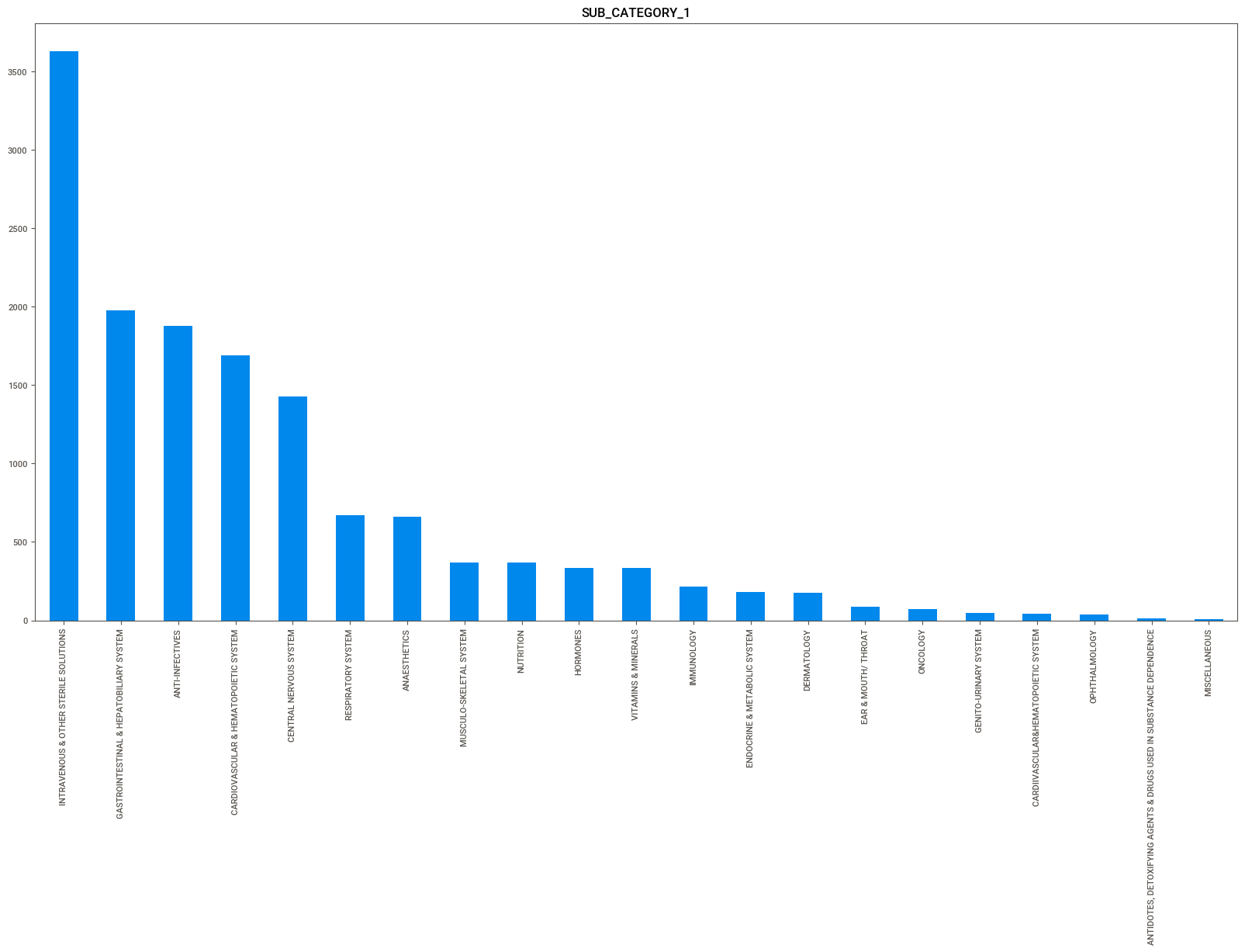
This bar chart represents least frequently used 20 drugs

**Subcat column:**

****

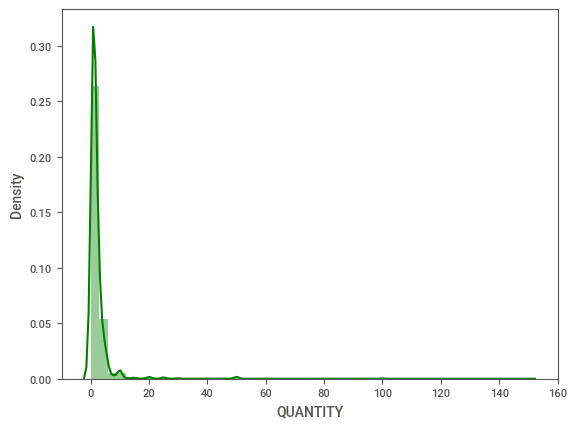
* There are 17 sub categories
* These sub categories include the type of drugs like injection, capsules and so on.
* Based on frequencies we conclude that injections are more contributing, followed by IV Fluids Electrolytes, TPN.
* Drops, vaccine, spray, patch, lotions, solution and sachets are least contributing.

**Subcat1 column:**

****

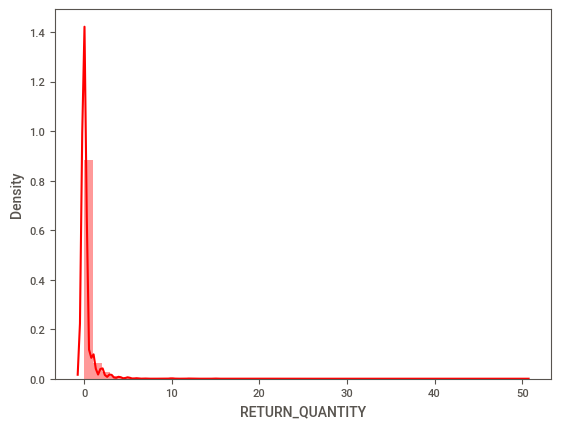
* There are 21 sub categories-1
* These sub categories-1 include the type of conditions like intravenous & other sterile solutions, Anti-infectives and so on.
* Based on frequencies we conclude that intravenous & other sterile solutions are more contributing, followed by Gastrointestinal & Hepatobiliary system and so on.
* Antidotes, Deoxifying agents & Drugs used in substance dependence are least contributing.

**Quantity Column:**

****

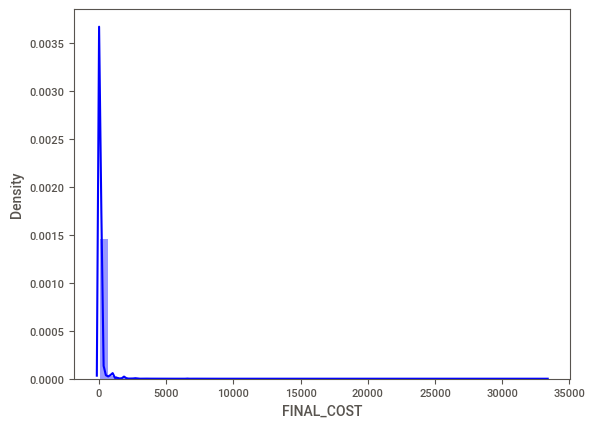
* The distribution is positively skewed or right skewed
* The data points are more contributing between 0 to 10 (approximately).
* So, we can say that patients frequently purchasing 1 to 10 quantities of drugs.
* The data points are less contributing more between 11 to 150(approximately).
* We can say that patients are not buying more quantity.

**Returned Quantity column:**

****

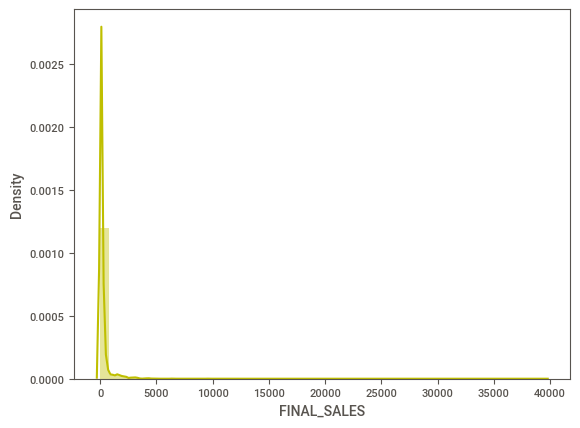
* The distribution is positively skewed or right skewed.
* The data points are more contributing between 0 to 2 (approximately).
* The return doesn’t happen most of the time.
* We can say that patients returning 1 to 2 quantities of drugs.
* The data points are less contributing between 3 to 22 (approximately).

**Final Cost Column:**

****

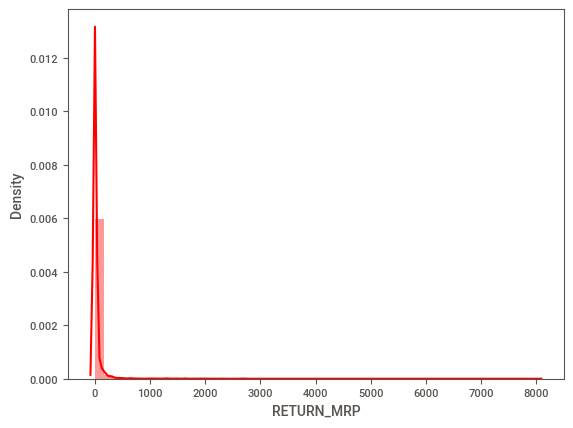
* The distribution is positively skewed or right skewed.
* The data points are more contributing between 0 to 1000 (approximately).
* The data points are less contributing between 1001 to 33000 (approximately).

**Final Sales Column:**

****

* The distribution is positively skewed or right skewed.
* The data points are more contributing between 0 to 1000 (approximately).
* The data points are less contributing between 1001 to 40000 (approximately).

**Return MRP Column:**

****

* The distribution is positively skewed or right skewed.
* The data points are more contributing between 0 to 100 (approximately).
* The data points are less contributing between 101 to 8100 (approximately).