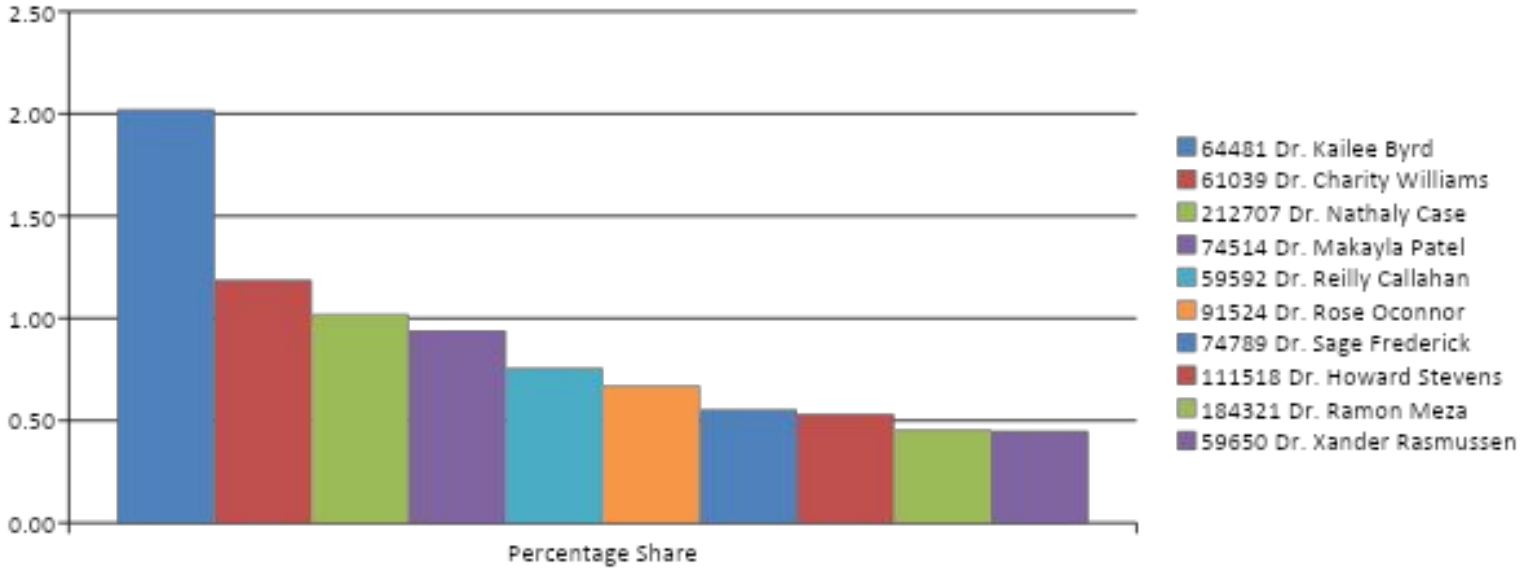
Data Harmonization

Exploring the data and its structure remains crucial before starting any analysis and I have explored and validated the data under the following two criteria:

1. Validation of data types, missing value analysis, data relationship , etc.
2. Business rule based discrepancies(eg. Orders, Doctors and Month details mappings, other features validations, etc)

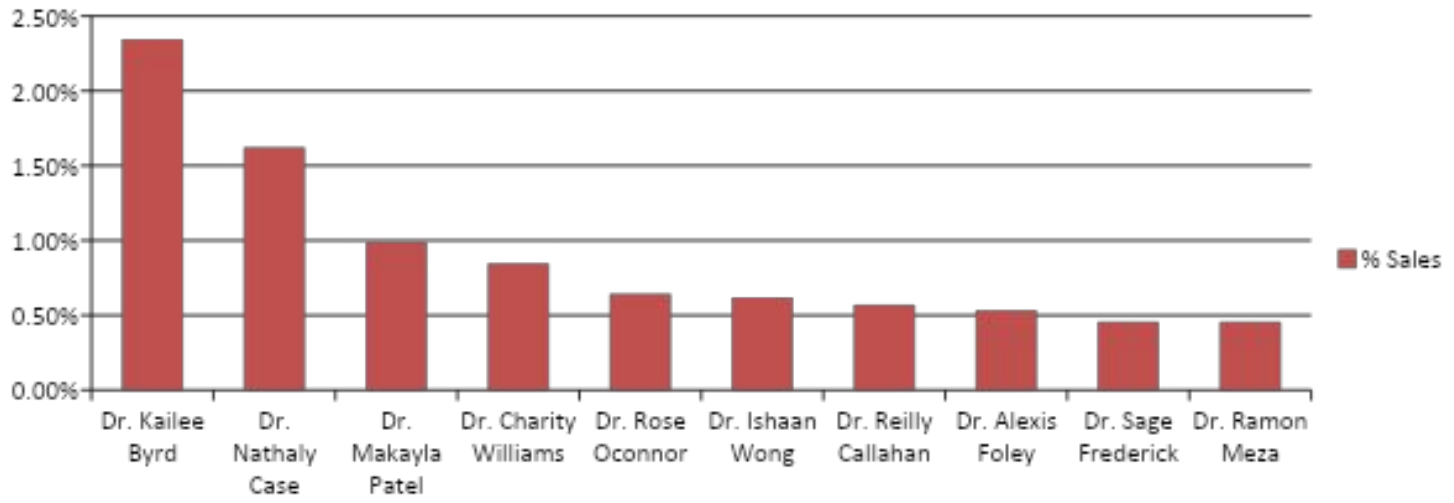
Highest Sales in 2019

AccountNumber	ContactName	Percentage Share
64481	Dr. Kailee Byrd	2.02%
61039	Dr. Charity Williams	1.19%
212707	Dr. Nathaly Case	1.02%
74514	Dr. Makayla Patel	0.94%
59592	Dr. Reilly Callahan	0.76%
91524	Dr. Rose Oconnor	0.67%
74789	Dr. Sage Frederick	0.55%
111518	Dr. Howard Stevens	0.53%
184321	Dr. Ramon Meza	0.46%
59650	Dr. Xander Rasmussen	0.45%



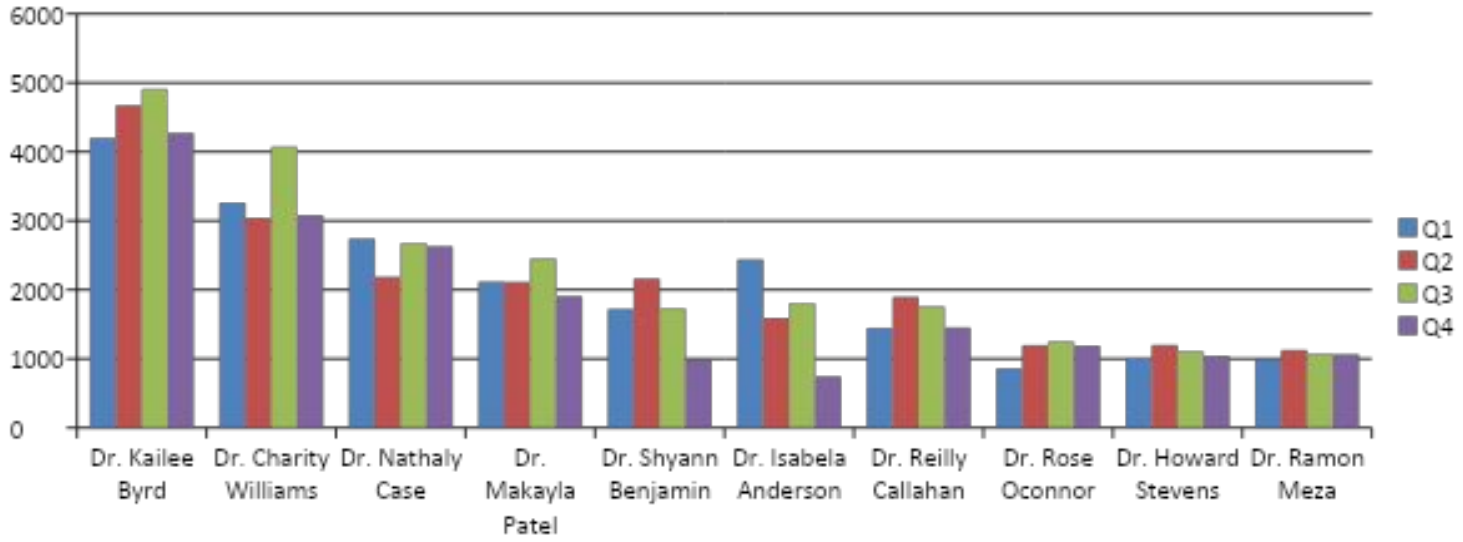
Highest Sales in Dec 2019

AccountNumber	ContactName	01-12-2020 Sales	% Sales
64481	Dr. Kailee Byrd	700	2.34%
212707	Dr. Nathaly Case	484	1.62%
74514	Dr. Makayla Patel	295	0.99%
61039	Dr. Charity Williams	253	0.85%
91524	Dr. Rose Oconnor	192	0.64%
64700	Dr. Ishaan Wong	184	0.62%
59592	Dr. Reilly Callahan	170	0.57%
238827	Dr. Alexis Foley	159	0.53%
74789	Dr. Sage Frederick	136	0.46%
184321	Dr. Ramon Meza	136	0.46%



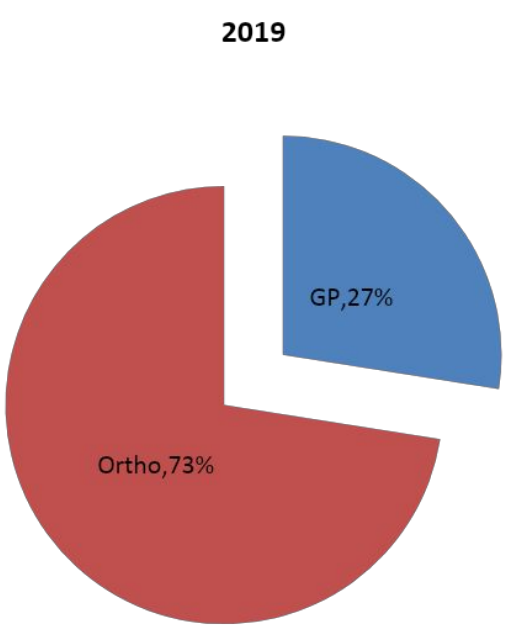
Top 10 Highest Sales

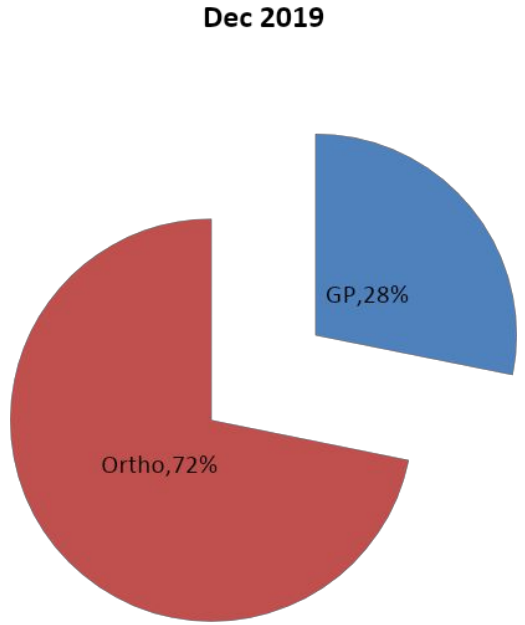
Doc Name	Q1	Q2	Q3	Q4	Grand Total
Dr. Kailee Byrd	4200	4670	4910	4270	18050
Dr. Charity Williams	3266	3036	4071	3082	13455
Dr. Nathaly Case	2740	2184	2672	2628	10224
Dr. Makayla Patel	2115	2110	2450	1905	8580
Dr. Shyann Benjamin	1719	2160	1728	981	6588
Dr. Isabela Anderson	2438	1590	1802	742	6572
Dr. Reilly Callahan	1445	1895	1755	1450	6545
Dr. Rose Oconnor	858	1194	1248	1188	4488
Dr. Howard Stevens	1008	1197	1106	1036	4347
Dr. Ramon Meza	994	1124	1066	1062	4246



Professional Category Wise Sales

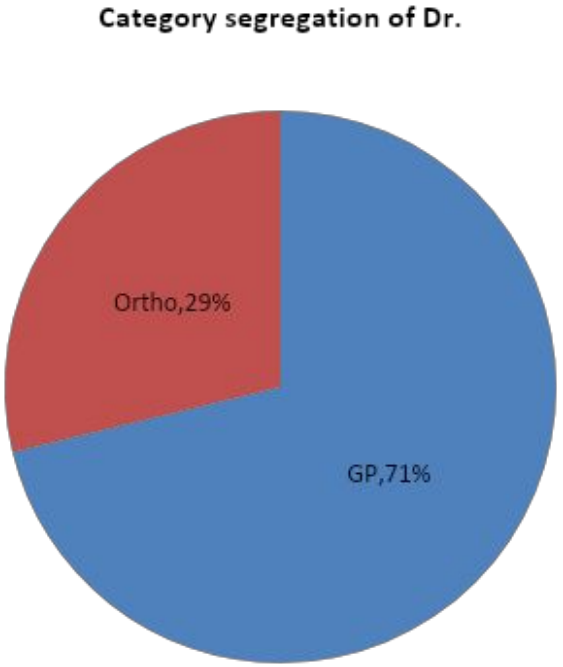
Year 2019			Month Dec 2019	
Professional Category	**Sum of Order**		**Professional Category**	**Sum of Order**
GP	246889		GP	8413
Ortho	650970		Ortho	21443





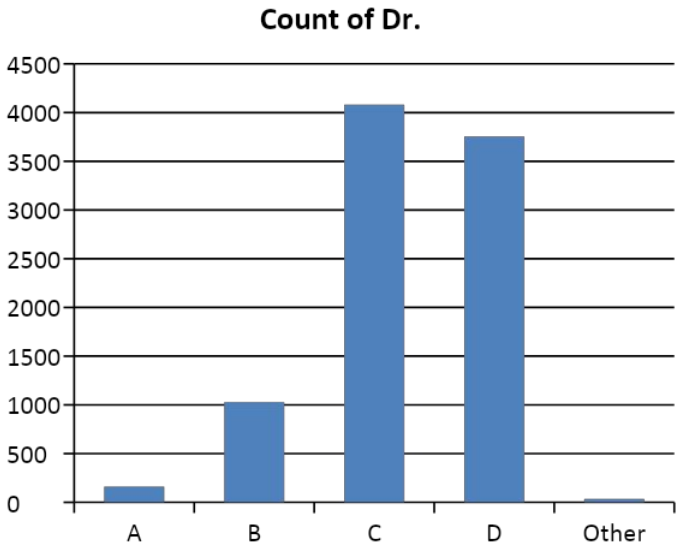
Category wise Dr. Segregation

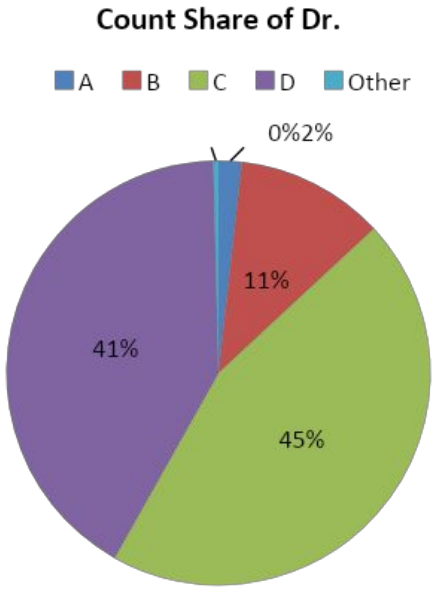
Row Labels	Count of ContactName
GP	6449
Ortho	2609



Segment wise Doctors Distribution

Row Labels	Count of ContactName
A	160
B	1027
C	4081
D	3755
Other	35

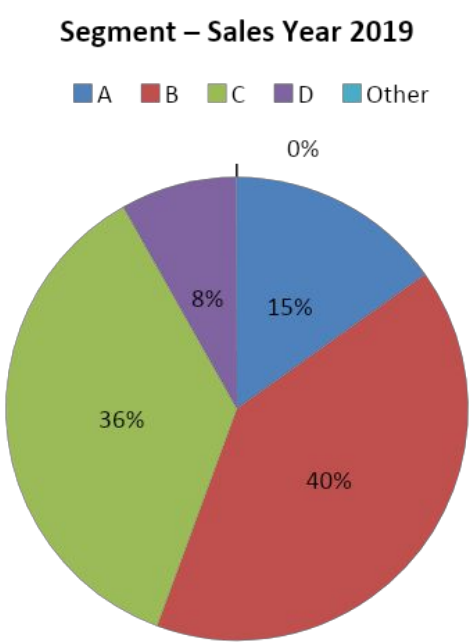


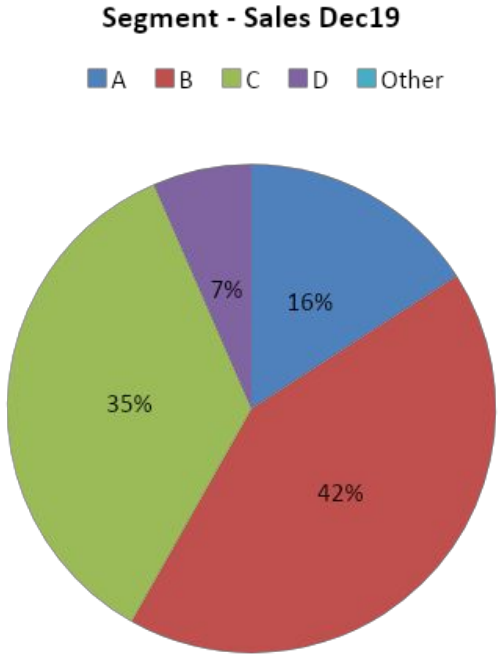


Sales distribution with respect Doctor's Segments

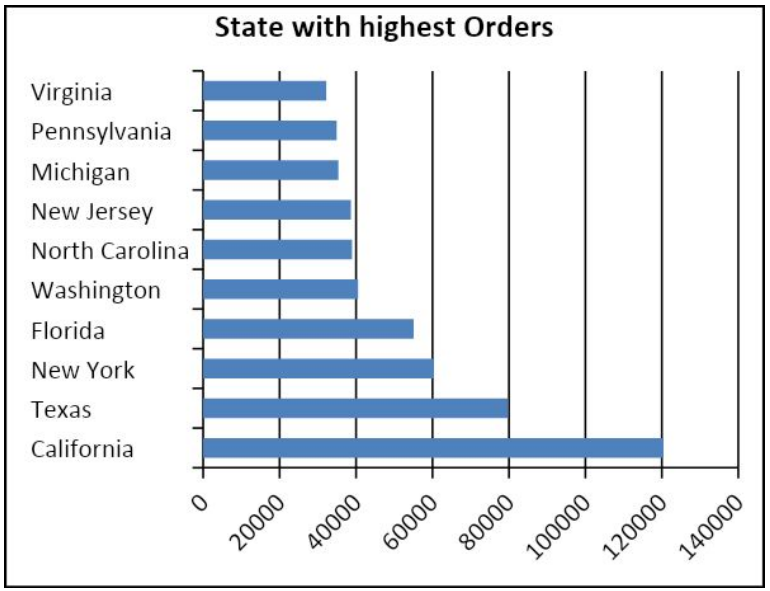
	2019 Sales				
Year	A	B	C	D	Other
2019	49782	133206	119125	26796	110

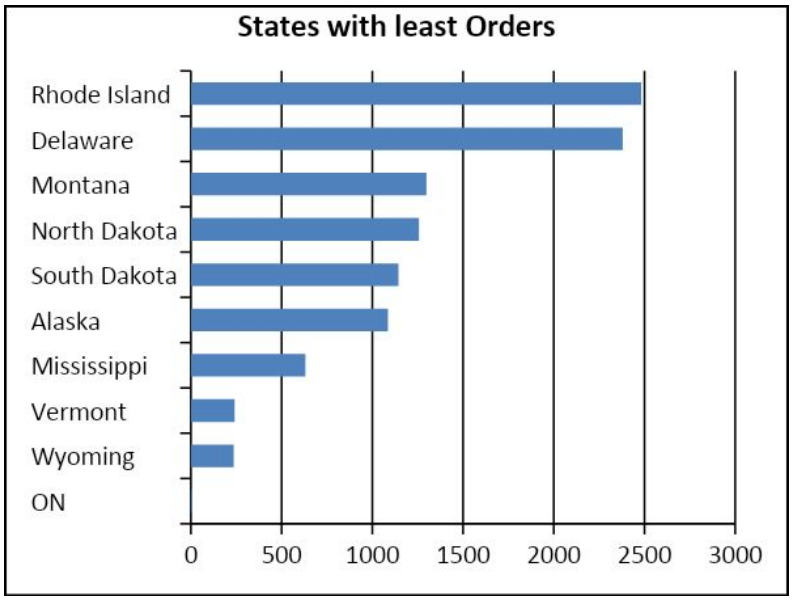
	Dec 2019 Sales				
Month	A	B	C	D	Other
Dec-19	4739	12627	10536	1954	0

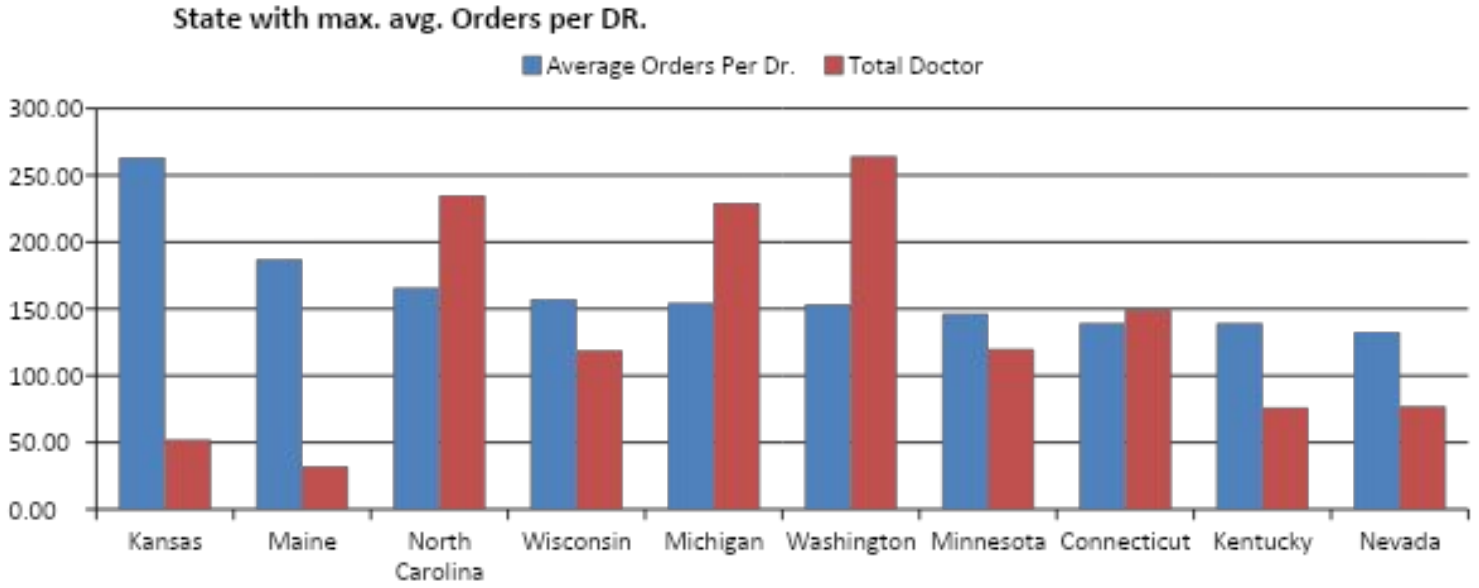




State wise analysis

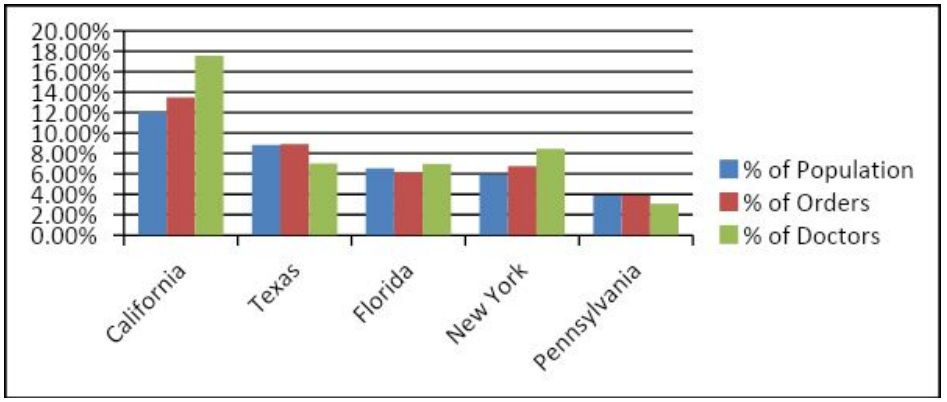






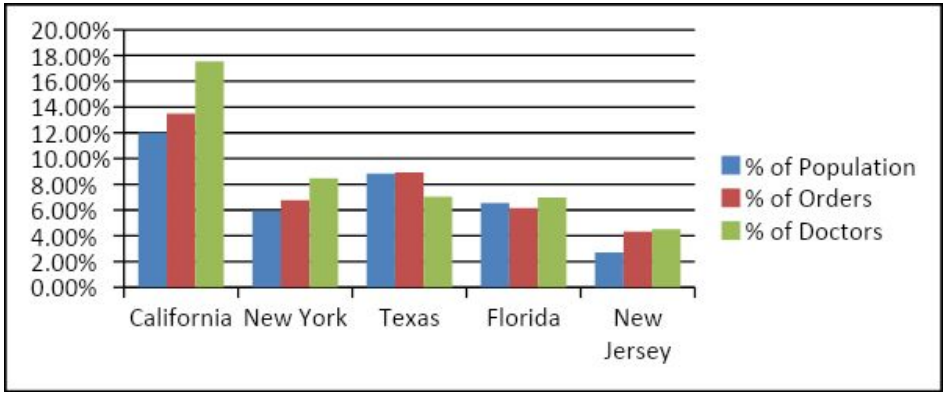
State Population wise Distribution (Top States)

State	% of Population	% of Orders	% of Doctors
California	12.04%	13.48%	17.55%
Texas	8.83%	8.92%	7.02%
Florida	6.54%	6.16%	6.98%
New York	5.93%	6.75%	8.45%
Pennsylvania	3.90%	3.91%	3.08%



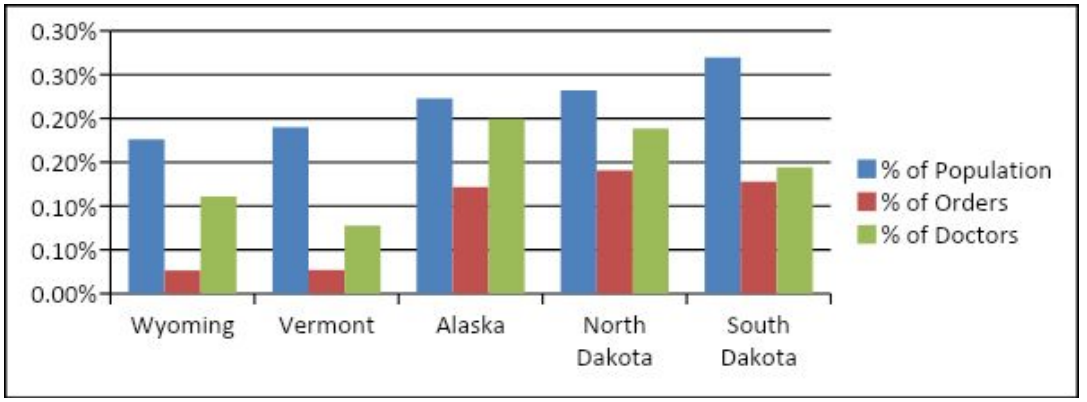
State	% of Population	% of Orders	% of Doctors
California	12.04%	13.48%	17.55%
Texas	8.83%	8.92%	7.02%
New York	5.93%	6.75%	8.45%
Florida	6.54%	6.16%	6.98%
Washington DC	2.53%	4.53%	2.93%

State	% of Population	% of Orders	% of Doctors
California	12.04%	13.48%	17.55%
New York	5.93%	6.75%	8.45%
Texas	8.83%	8.92%	7.02%
Florida	6.54%	6.16%	6.98%
New Jersey	2.71%	4.33%	4.51%

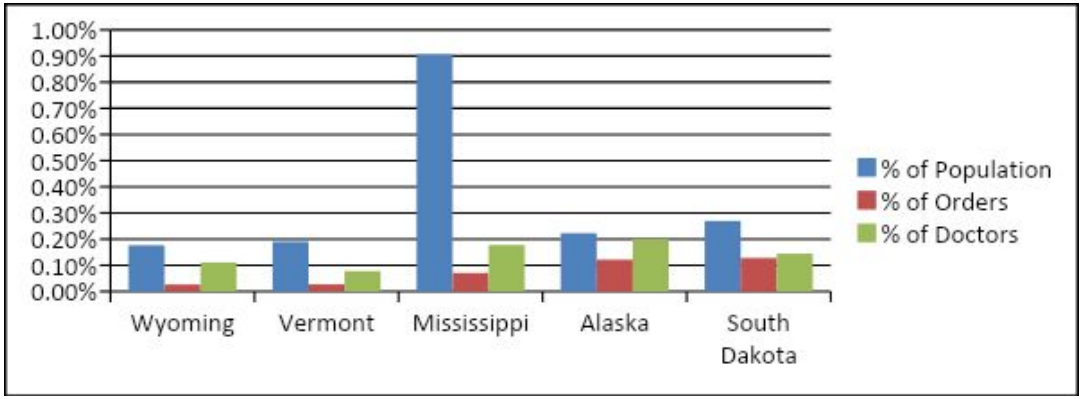


State Population wise Distribution (Bottom States)

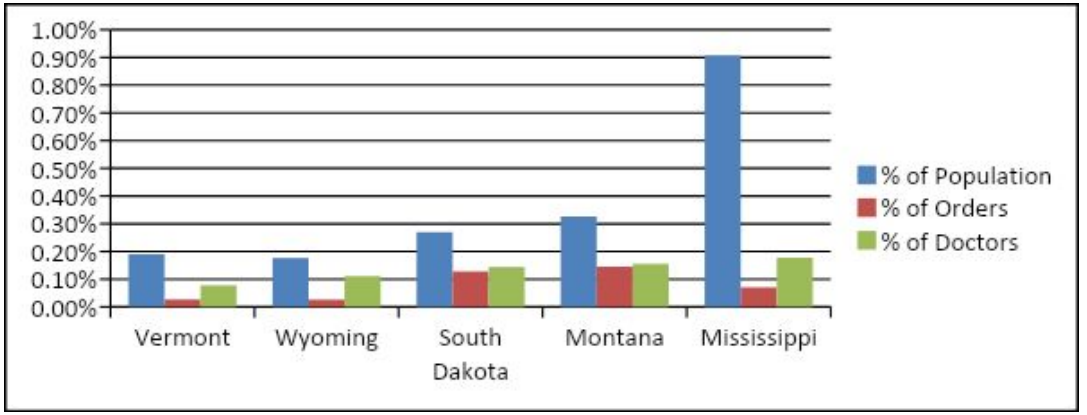
State	% of Population	% of Orders	% of Doctors
Wyoming	0.18%	0.03%	0.11%
Vermont	0.19%	0.03%	0.08%
Alaska	0.22%	0.12%	0.20%
North Dakota	0.23%	0.14%	0.19%
South Dakota	0.27%	0.13%	0.14%



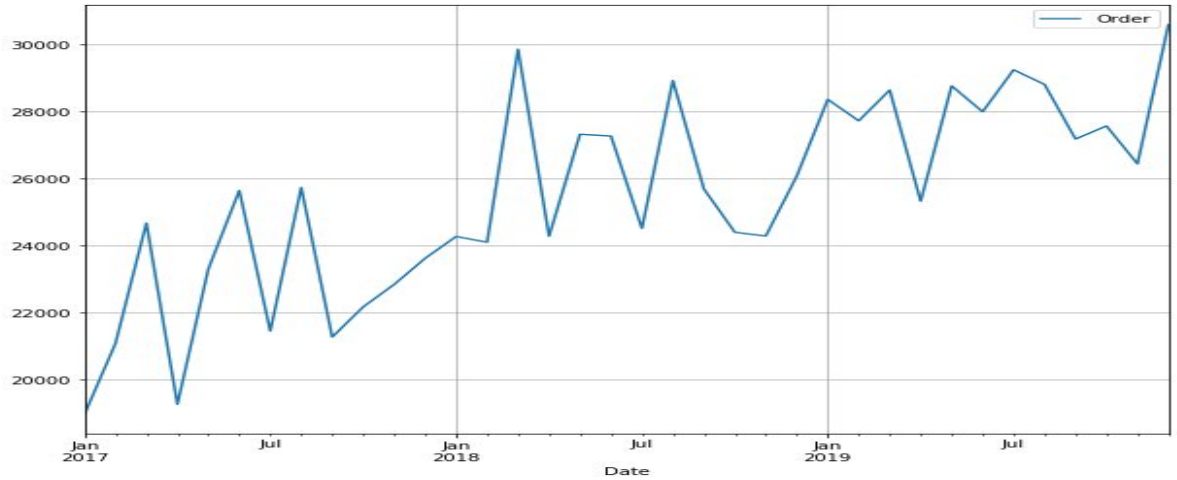
State	% of Population	% of Orders	% of Doctors
Wyoming	0.18%	0.03%	0.11%
Vermont	0.19%	0.03%	0.08%
Mississippi	0.91%	0.07%	0.18%
Alaska	0.22%	0.12%	0.20%
South Dakota	0.27%	0.13%	0.14%

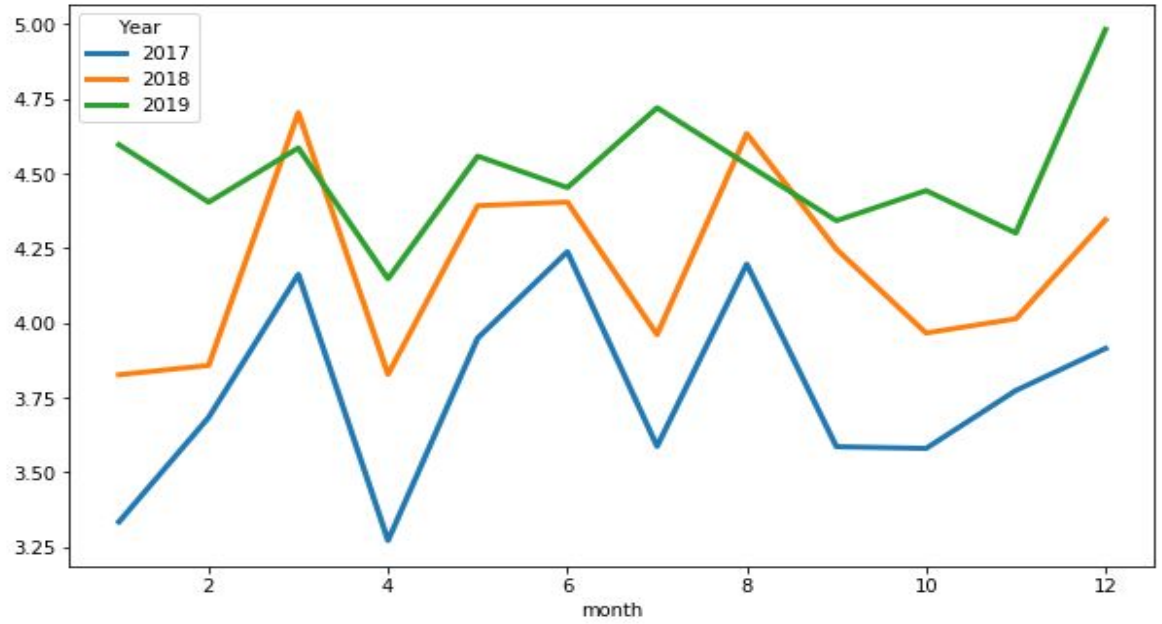


State	% of Population	% of Orders	% of Doctors
Vermont	0.19%	0.03%	0.08%
Wyoming	0.18%	0.03%	0.11%
South Dakota	0.27%	0.13%	0.14%
Montana	0.33%	0.15%	0.16%
Mississippi	0.91%	0.07%	0.18%



Trends and Seasonality in Data set





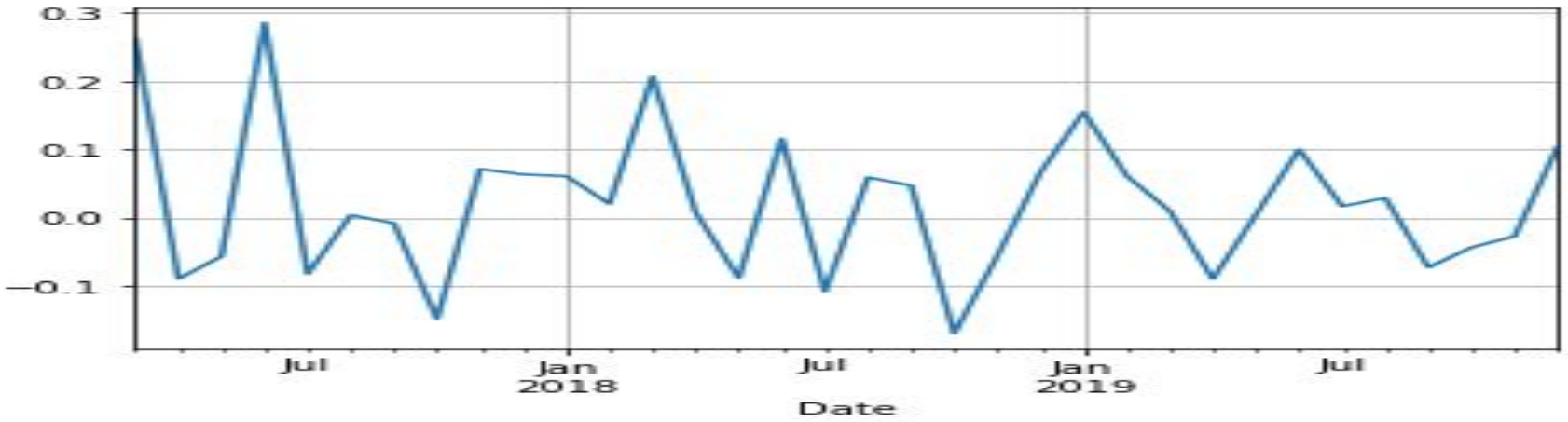
The total sales clearly shows **UPWARDS** trends as the total sales kept on increasing yearly.

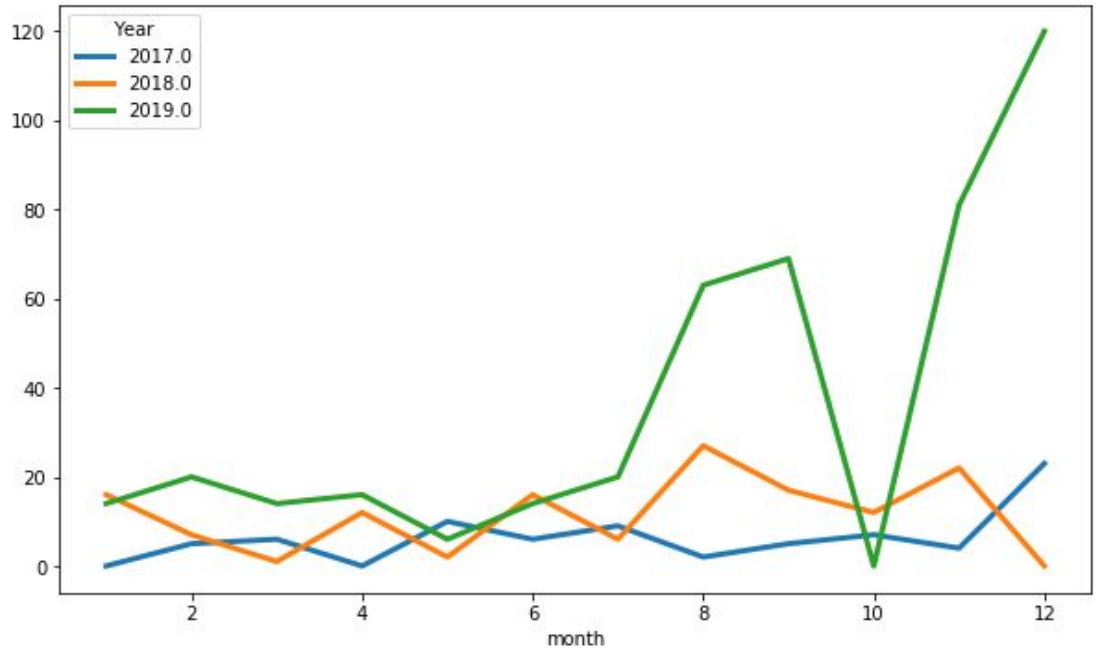
Graph also depicts some **SEASIONALITY** trends, as we can see that:

- There is always a fall in total sales at the beginning of the year.
- Also the sales increases in Feb and again decreases

```
ADF Statistic: -1.395508546974049
p-value: 0.5844274732972616
Critical Values:
        1%: -3.6996079738860943
        5%: -2.9764303469999494
        10%: -2.627601001371742
```

Stationary data in overall sales





ADF Statistic: -3.1037866464157005

p-value: 0.02627422587566269

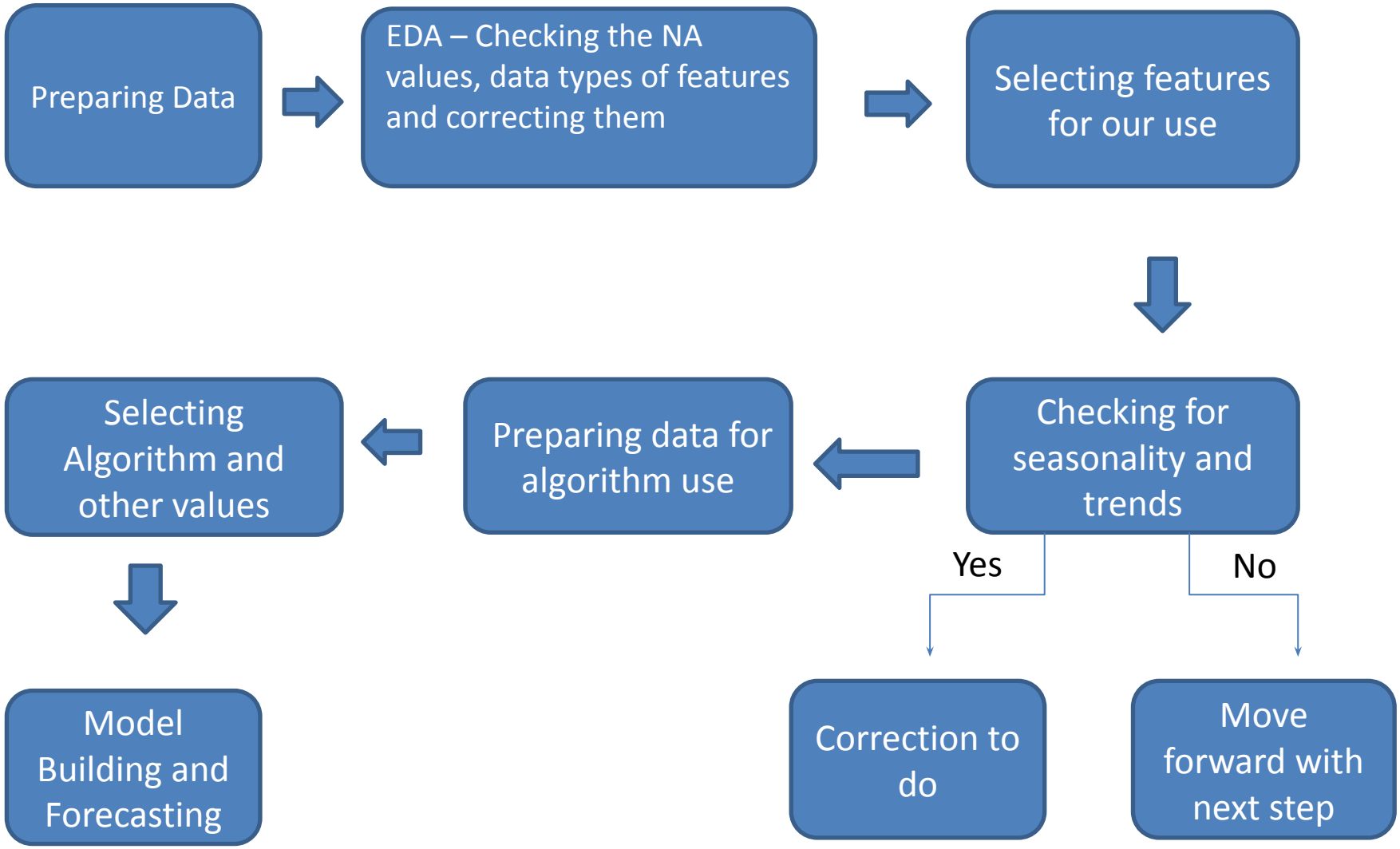
Critical Values:

1%: -3.7377092158564813

5%: -2.9922162731481485

10%: -2.635746736111111

Forecasting Steps



Preparing Data and EDA

1. Inserting Libraries
2. Importing data sets (both sheets)
3. Merging both the datasets
4. Removing Blank spaces doing other EDA
5. Checking the data types of features

Feature Selection and seasonality and trend test

1. Selection of Datetime feature and the dependent feature
2. Correcting the dataset if required.
3. Mean and variance method used for checking the seasonality.
4. Different graphs used for viewing the seasonality.

Correction of Seasonality

1. ADF and KPSS are used for checking if the data is stationary or not.
2. Converting non stationary data into Stationary data using log shift technique.

Selecting Algorithm, Model Building and Forecasting

1. Checking the autocorrelation and partial autocorrelation (value of p and q)
2. Building model with ARIMA and ARMA and checking the errors
3. Building the different models using different order values.
4. Selecting the model with the optimum errors
5. Doing forecasting using the selected model
6. Backtracking the actual value.

Strengths and Weakness of the Model

Strengths

1. Very less mean square error at 0.0053 for monthly forecasting, hence we get quite accurate error.
2. Simple to understand.

Weakness

1. Approximation is done in removing the seasonality and trends.
2. Approximation is done while converting the non stationary data into stationary.

Findings, Analysis and Insights

1. Dr. Kailee Byrd has the highest no. of orders in the year 2019, also he has the highest no. of orders for Dec 2019.
2. Dr. Charity Williams has the 2nd highest orders in 2019 but his order count decreases in Dec 2019.
3. Dr. Kailee Byrd and Dr. Charity Williams are very consistent in there order counts over different quarters.
4. As per the professional Category is concerned, Ortho has nearly 73% share in the orders in 2019 and 72% share in Dec 2019.
5. 29% of Dr. are from Ortho and 71% are from GP.
6. 4081 i.e. 45% of Dr. are in Segment C and this segment has the highest count where as only 35 Dr. are in other segment which is lowest.
7. California produces the highest no. of Orders and Wyoming has the least orders.
8. Kansas has the maximum average orders per Doctors, which is nearly 250 orders per Dr.
9. Population wise California has the highest population share at 12.04%, also California has 13.48% of orders and also it has the highest percentage of Doctors count.
10. Wyoming has the least population share at 0.15%, and also has the lease Order share where as Vermont has least Doctor share.

Possible Recommendations

1. We have opportunity in Mississippi as the Population vs Doctor ratio is very high by assigning more Doctors there.
2. Kansas, Maine, North Carolina are the states where Average Orders per Doctors is very High, So more Doctors can be assigned to these states to share the work loads and states like Wyoming, Vermont, Mississippi are the states with least average orders per Doctors hence few Doctors can be shifted to other states.