Pharmaceutical Sales Analysis

Victoria Tamburro

Data Collection

- Collected and groomed by researcher Milan Zdravković.
- The initial dataset was created from transactional sales data collected from a single pharmacy's point-of-sales system.
- The 57 drugs sold were then grouped to the Anatomical Therapeutic Chemical (ATC) Classification System Categories.
  - M01AB - Anti-inflammatory and antirheumatic products, non-steroids, Acetic acid derivatives and related substances
  - M01AE - Anti-inflammatory and antirheumatic products, non-steroids, Propionic acid derivatives
  - N02BA - Other analgesics and antipyretics, Salicylic acid and derivatives
  - N02BE/B - Other analgesics and antipyretics, Pyrazolones and Anilides
  - N05B - Psycholeptics drugs, Anxiolytic drugs
  - N05C - Psycholeptics drugs, Hypnotics and sedatives drugs
  - R03 - Drugs for obstructive airway diseases
  - R06 - Antihistamines for systemic use
- The data was collected for six years, with an incomplete final year.

Contents

Introduction

- Through this analysis data trends will be discovered and forecasting models will be created to predict long term generalized trends for specific drug groups.
- These machine learning forecasting models and seasonality trends can be used to inform strategic re-ordering.
- Recommendations will be made for further analysis based on exogeneous variables.

Understanding the Data

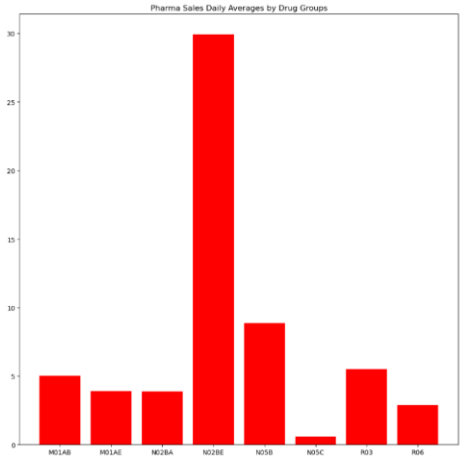
- The data is presented in several views which include hourly, daily, weekly, and monthly point-in-time sales.
- For the purposes of this analysis, we assume that the values presented represent the number prescriptions filled.

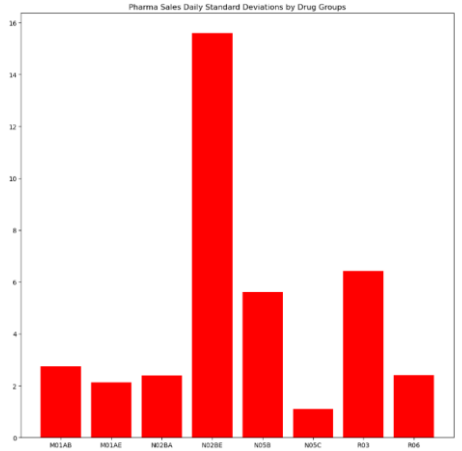
Example Data Set:

datum	M01AB	M01AE	N02BA	N02BE	N05B	N05C	R03	R06	Year	Month	Hour	Weekday Name
1/2/2014	0.0	3.67	3.4	32.40	7.0	0.0	0.0	2.0	2014	1	248	Thursday
1/3/2014	8.0	4.00	4.4	50.60	16.0	0.0	20.0	4.0	2014	1	276	Friday
1/4/2014	2.0	1.00	6.5	61.85	10.0	0.0	9.0	1.0	2014	1	276	Saturday
1/5/2014	4.0	3.00	7.0	41.10	8.0	0.0	3.0	0.0	2014	1	276	Sunday
1/6/2014	5.0	1.00	4.5	21.70	16.0	2.0	6.0	2.0	2014	1	276	Monday

Descriptive Statistics Highlights

- Upon inspection of the descriptive statistics it can be noted that the daily, weekly, and monthly collected data increases proportionately for the mean, median, standard deviation.
- It as also evident that the drug group N02BE is the most sold drug, and the drug group N05C is the least sold drug.
- There is also a high level of variance for the drug group N02BE, alluding to its seasonal influence which will be covered in later slides.





Descriptive Statistics

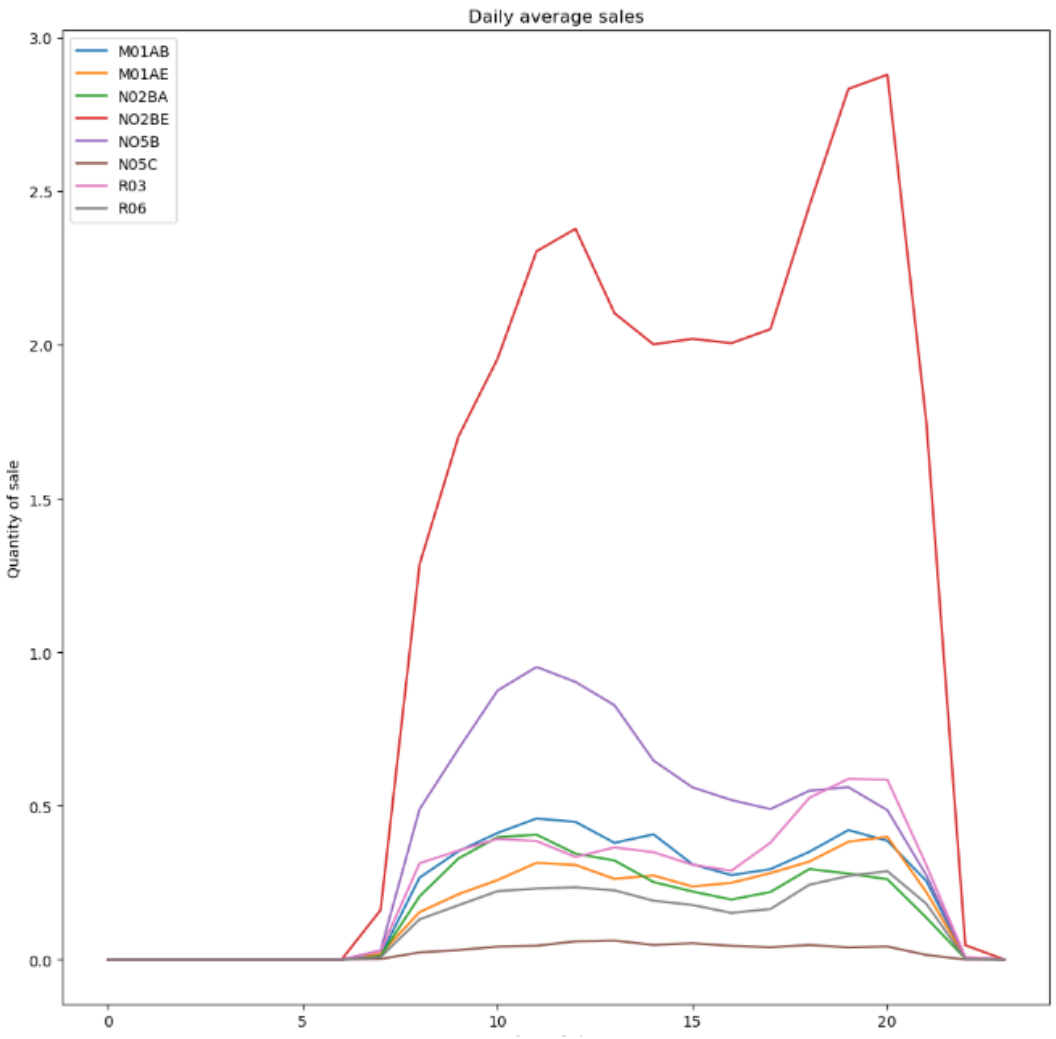
	M01AB	M01AE	N02BA	N02BE	N05B	N05C	R03	R06
count	2106	2106	2106	2106	2106	2106	2106	2106
mean	5.033683	3.89583	3.880441	29.917095	8.853627	0.593522	5.512262	2.900198
std	2.737579	2.133337	2.38401	15.590966	5.605605	1.092988	6.428736	2.415816
min	0	0	0	0	0	0	0	0
25%	3	2.34	2	19	5	0	1	1
50%	4.99	3.67	3.5	26.9	8	0	4	2
75%	6.67	5.138	5.2	38.3	12	1	8	4
max	17.34	14.463	16	161	54.833333	9	45	15

M01AB	M01AE	N02BA	N02BE	N05B	N05C	R03	R06
302	302	302	302	302	302	302	302
35.102441	27.167611	27.060295	208.627161	208.627161	4.138935	38.439811	20.224561
8.617106	7.043491	8.086458	76.069221	76.069221	3.129265	22.900873	11.381464
7.67	6.237	3.5	86.25	86.25	0	2	1
29.3875	22.3875	21.3	149.3	149.3	2	21	11.475
34.565	26.7895	26.5	198.3	198.3	3.979167	35	17.5
40.175	31.0465	32.475	252.4715	252.4715	6	51	26
65.33	53.571	60.125	546.899	546.899	17	131	65

count	70	70	70	70	70	70	70	70
mean	149.992	116.514286	115.020843	892.542071	262.118571	17.842857	167.675	86.662571
std	31.485325	27.889336	31.245899	338.843908	85.06093	8.481242	81.767979	45.859336
min	0	0	0	0	1	0	0	0
25%	137.49	103.51825	94.375	648.1875	223.75	12	112	49.875
50%	154.635	114.84	117.225	865.8245	250.3	18	160	74.1
75%	169	128.35975	133.8375	1061.58	293.65	23	218.25	119.8075
max	211.13	222.351	191.6	1856.815	492	50	386	213.04

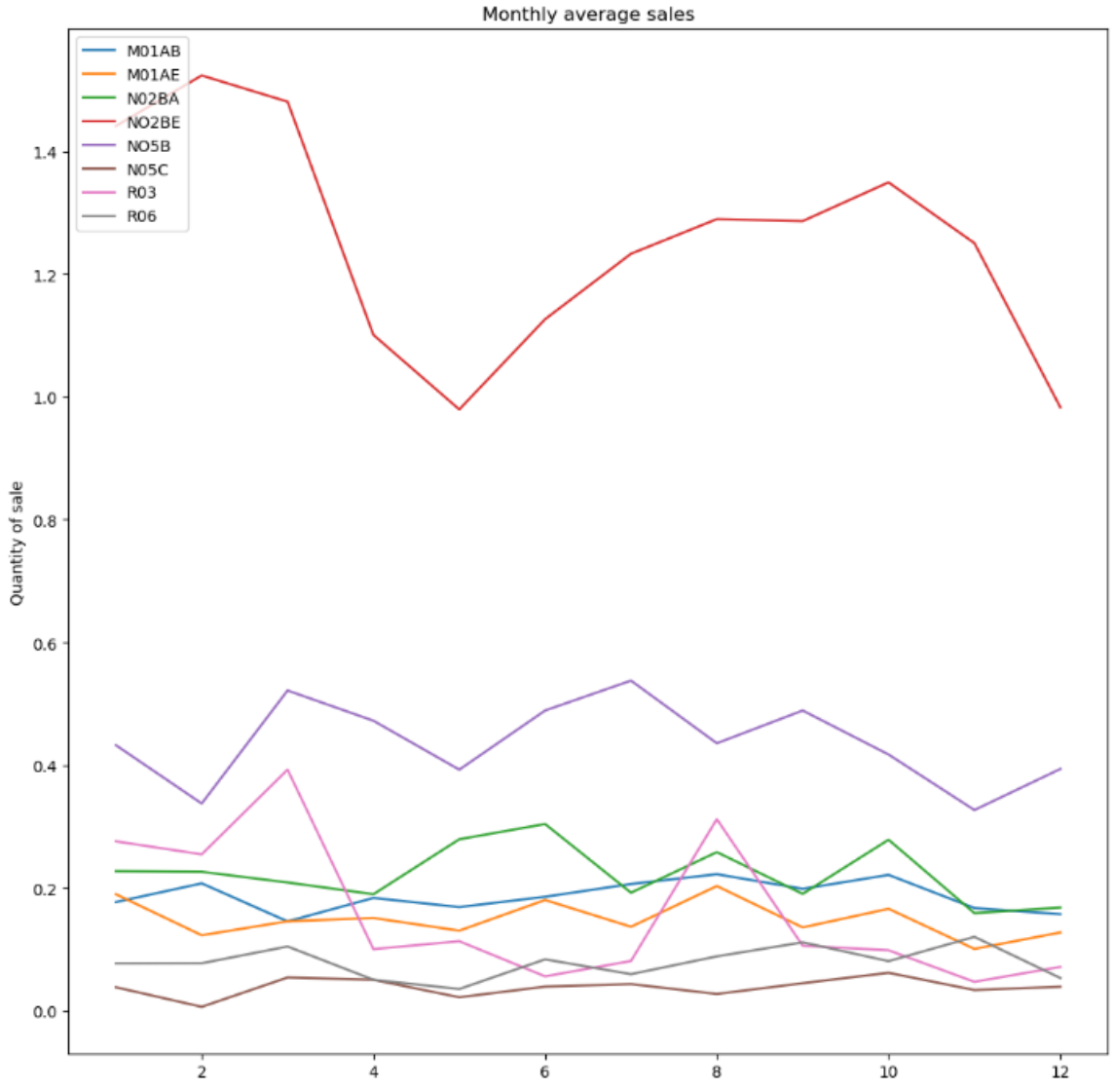
Time of Day

- The data was aggregated by average sales by time of day. Upon visualization it can be observed that on average the flux of drugs sold at various times of the day seems to differ based on these groups.
- Notably, N02BE has two peaks at 12:00PM, and 9:00PM with a steep drop in purchase frequency, while R03 has only an evening peak which has a more fluent onset.
- This could be perhaps due to the nature of the drug and the availability of the patients that are able to visit the store during those hours. The times seem to trend with a lunch time and evening time store visit.
- Given further customer demographic information and customer analysis, this could impact certain aspects of the business such as store staffing and shipment/loading scheduling.



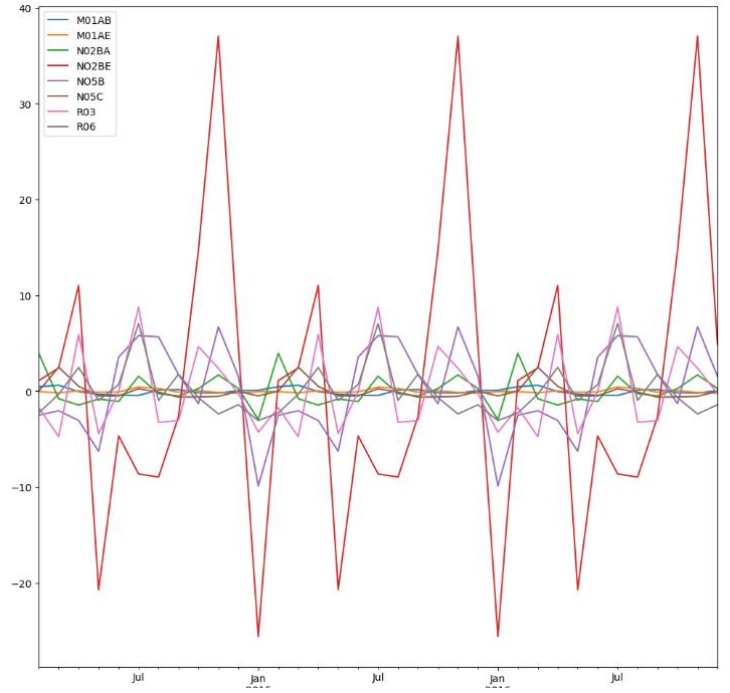
Time of Month

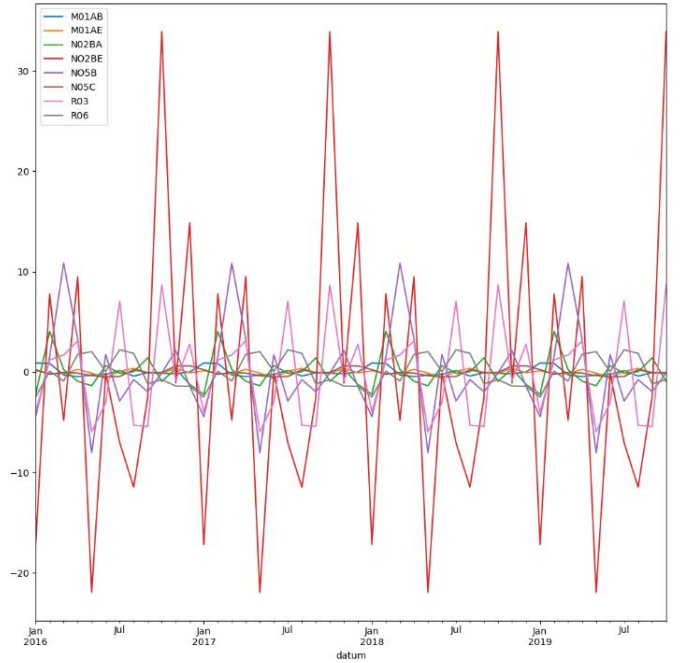
- Additionally, the data was aggregated by average sales by time of month. Several trends can be seen.
- N02BE has a sharp spike in February and a smaller spike in October. This is an interesting trend as these drugs are generally painkillers, which one wouldn't assume to have any type of seasonality, such as R06 which is an antihistamine for seasonal allergies.
- Seasonal trends can be observed for R06 aligning with Spring, Summer and Fall allergy seasons.
- With further data collection on patient diagnoses it may be worth investigating further into the reason for N02BE trending as it is the highest selling drug of the pharmacy.



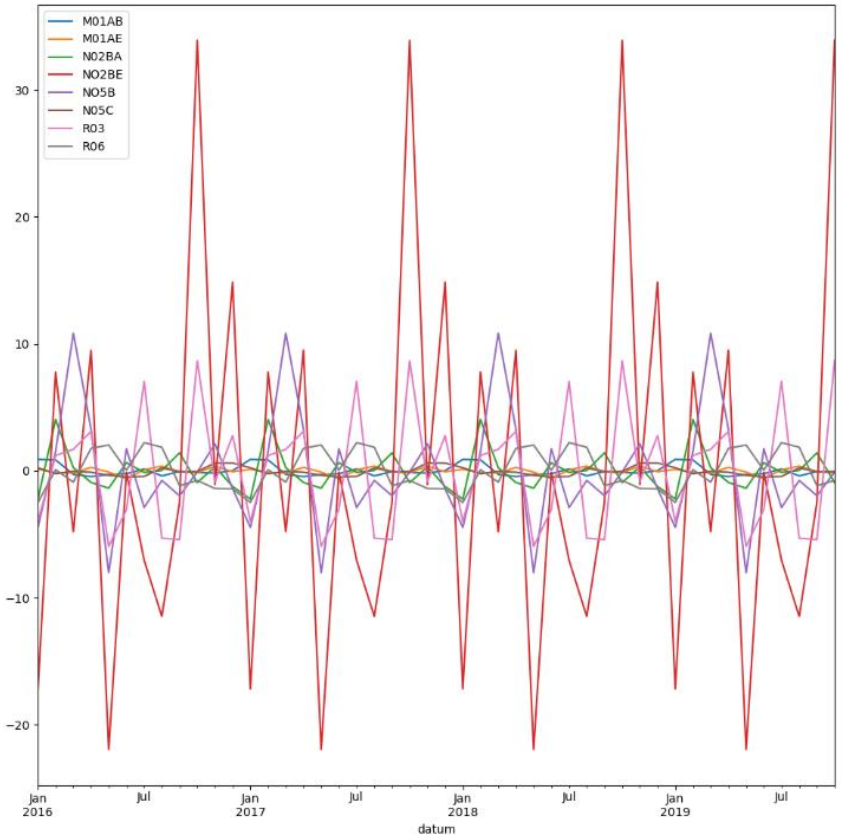
Seasonality

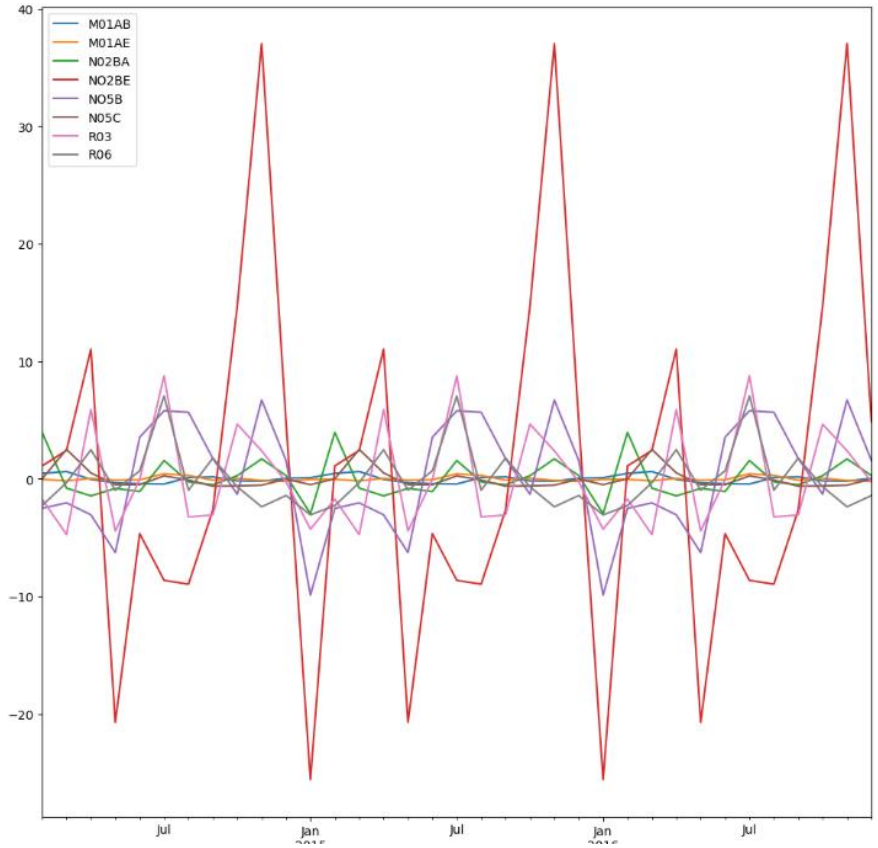
- The data was split into three-year intervals for clarity of visualization. The data was then decomposed by drug group using an additive model, to indicate the impact of the seasonality.
- It can be observed that the impact of the seasons in the time series dataset is the greatest on NO2BE





Seasonality Detailed

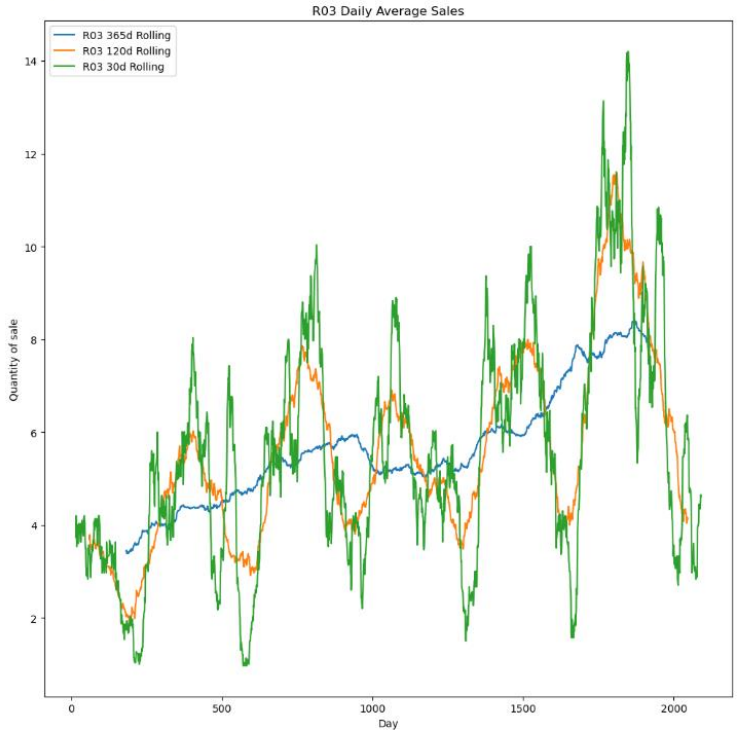




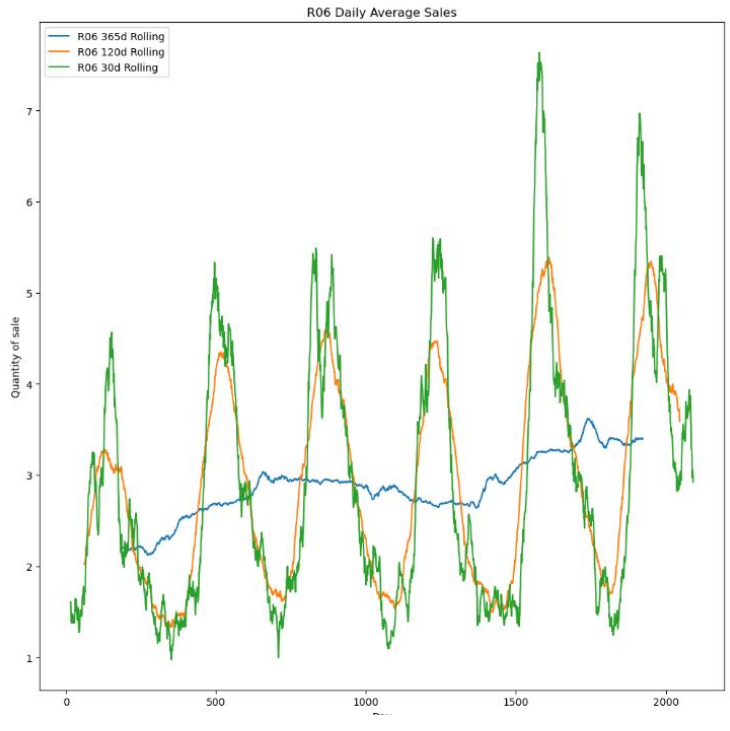
Observing Long-Term Trends

- Because of the volatile sales volume changes due to seasonality, it was required to normalize the data with moving averages to get a scope of the long-term drug sale trends.
- The following visualizations were created from 30, 120, and 365 day intervallic moving averages. A linear regression model was performed on the 365-day normalized averages, and three of the drugs displayed clear linear tendencies over the six years: R03, R06, and N02BA.
- These trends can be indicative of effectiveness, comfort of prescription from medical providers, advertising, and potential endemic/pandemic tendencies.

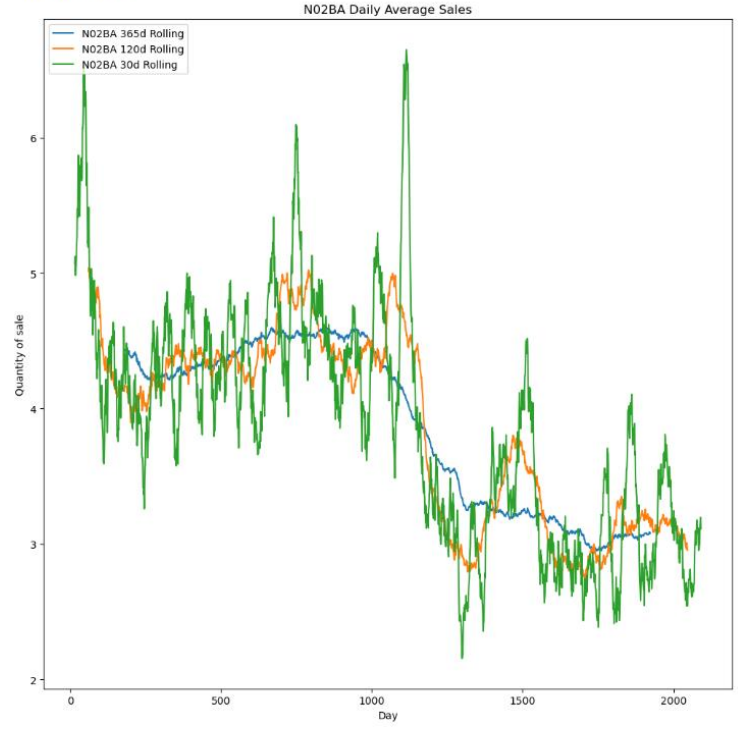
Training set score: 0.83
Test set score: 0.81

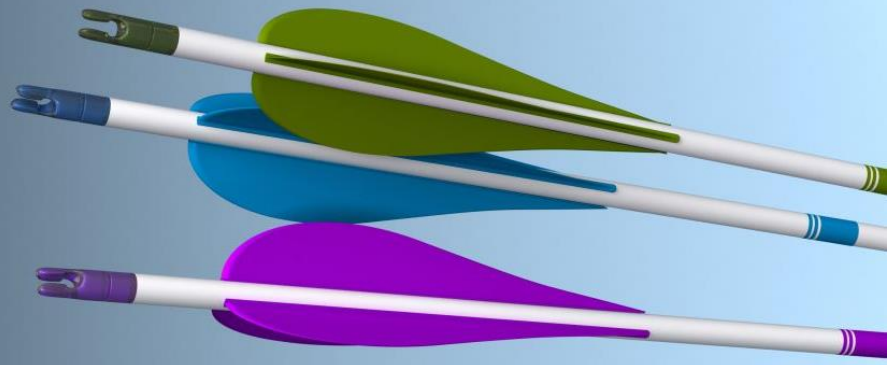


Training set score: 0.70
Test set score: 0.66



Training set score: 0.76
Test set score: 0.76





Linear Regressions

- N02BA Predictive Model - Reducing over Time
  - 76% Accuracy Training Score
  - 76% Accuracy Test Score
- R03 Predictive Model – Increasing Significantly
  - 83% Accuracy Training Score
  - 81% Accuracy Test Score
- R06 Predictive Model – Increasing Slightly
  - 70% Accuracy Training Score
  - 66% Accuracy Test Score

Conclusion

- The forecasting models can be used to predict long term generalized trends for specific drug groups. These forecasting models and seasonality trends can inform strategic re-ordering of these drugs.
- It should be noted that while these predictions show the trends of sales, there are other factors that should also play a part in the re-ordering of a drug.
  - Pharma Capacities - Shelf/Stocking space availability
  - Price Optimizations
  - Essential Rescue drugs that must be kept on hand such as R03 which is used for obstructive airway diseases

References

- *Pharma sales data*. (n.d.). Www.kaggle.com. Retrieved December 18, 2022, from https://www.kaggle.com/datasets/milanzdravkovic/pharma-sales-data
- *statsmodels.tsa.seasonal.seasonal_decompose — statsmodels*. (n.d.). Www.statsmodels.org. https://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.seasonal_decompose.html
- *sklearn.linear_model.LogisticRegression — scikit-learn 0.21.2 documentation*. (2014). Scikit-Learn.org. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html