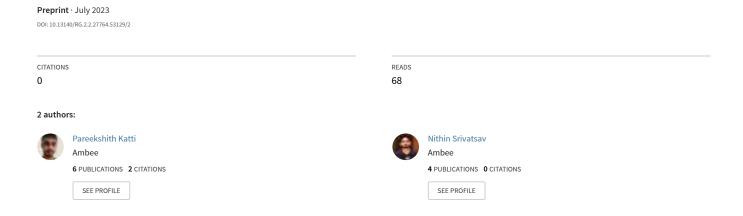
Analysis of the Role of Airborne Pollen in Seasonal Variation of Antihistamine Sales in Serbia



Analysis of the Role of Airborne Pollen in Seasonal Variation of Antihistamine Sales in Serbia

Pareekshith US Katti, N Nithin Srivatsav

Abstract

Airborne pollen is one of the leading causes of allergic rhinitis. In this study, we analyzed the relationship between pollen concentration and sales of antihistamines at a pharmacy in Niš, Serbia at daily, weekly, and monthly levels. Our analysis showed that pollen concentration had a significant impact on the sales of antihistamines at weekly and monthly levels. The sales peaked during the pollen seasons of Birch, Oak, and Pine trees. Multiple General Linear Models (GLMs) were built to model sales at daily, weekly, and monthly levels for both current sales and future sales. The best model was achieved when modeling the next month's total sales using the current month's total pollen. The model had an R2 score of 0.71. The result of this study may help the healthcare industry to make informed decisions to assess antihistamine demand using airborne pollen concentration.

Introduction

Allergic Rhinitis or Hay Fever is an allergic reaction commonly caused due to airborne pollen (Meltzer et al., 2009). Pollen levels are high during the spring and summer months which increases the chance of allergic rhinitis (Werchan and Bergmann, 2018)(Bitz et al., 2019). Antihistamine drugs are used to control allergic reactions triggered by our immune system (Randall and Hawkins, 2018), hence the sales

of the drugs might have a relation to pollen seasonal patterns. Understanding pollen seasonal patterns might lead to better management and sales of antihistamine drugs, helping both allergy sufferers and the pharmaceutical industry. (Guilbert et al., 2016) While pollen allergies are common, access to pollen data is limited, with most of the data not being public and freely available, therefore hindering both research and the progress of pollen-related studies. The availability of frequent pollen-level information globally is limited as traditional pollen-counting methods require mechanical equipment and manual labor. (Polling et.al., 2021) Various challenges such as the need for specialized and expensive equipment, experts, and microscopy have limited pollen counting to a very small set of locations (Buters et.al., 2018). Since the method involves manual counting of pollen, the data collection and publishing might be irregular and have a low temporal resolution (Buters et.al., 2022). The lack of standard ways to collect and count pollen data, different standards in determining risk as well as different formats for measuring and publishing pollen data have also made it difficult to expand research related to pollen allergies. (Comtois et.al 1999)

Airborne pollen concentration has significant influence by factors such as temperature and humidity along with other meteorological parameters (Bartková-Ščevková, 2003) (Peternel et.al., 2004) (Zewdie et.al, 2019). Recent advancements in fields such as machine learning have made it possible to estimate pollen to an extended set of locations using factors such as phenology, vegetation, weather parameters, and historical patterns (Zewdie et.al, 2019). Using machine learning, it is possible to lower the cost of collecting and counting pollen data by automating the process, standardizing methods of data estimation, and also reducing manual errors in counting. Machine learning has also helped in providing insights into historical patterns and future forecasts. (Notas et.al, 2015) (Csépe et.al., 2020)

Using modern machine learning based pollen datasets, it has become possible to analyze pollen patterns with other seasonal patterns that might be dependent on pollen seasonality such as antihistamine sales. In this study, we analyzed the seasonal patterns of pollen levels with seasonal patterns of antihistamine sales at a pharmacy in Niš, Serbia.

Literature Review

Pollen is a significant contributor to allergies in Europe. Hence it is also a driving factor when it comes to allergy medication sales. Several studies in Europe have found out pollen is one of the major contributors to antihistamine sales in Europe. A study conducted in Belgium for the period between 2005 - 2011 revealed that the medication sales had a significant impact from species such as Birch (Betula), Hornbeam (Carpinus), and Grasses (Gramineae). The study also noted that these species had high consequences for pollen allergies (Guilbert et al., 2016).

A similar study was conducted in the city of Cordoba from 1999 to 2001. The study noted that people from different parts of the city experienced different symptoms due to variances in location emissions. Major pollen species affecting the symptoms were Olive (Olae) and Grasses (Poaceae). The study also noted that the symptoms seemed to be worsening with time due to increased sensitization to urban tree species like plane (Sánchez-Mesa et.al., 2005).

A study conducted in New York from 2003 to 2008 found significant associations between allergy medications and tree pollen. The study also noted that the highest association was with 2-day logged tree pollen peak indicators. The study indicated that allergy medication sales had large increases following the peak of tree pollen season. The study also found that trees such as Oak, Birch, and Maple had significant predictive value for estimating allergic responses (Sheffield et.al., 2011). Another study conducted

in New York from 2002 to 2012 found that mid-spring pollen such as Oak, Birch, Maple, Ash, Beech, and Sycamore had the highest association with over-the-counter drug sales. The study also noted that the allergy symptoms were more in children (Ito et.al., 2015). In Ukraine, studies found that the sale of antihistamines showed an association with grass and tree pollen seasons. The study noted that the peaks started earlier for the period between 2015 to 2016 compared to 2008 to 2009. A new pattern of sales was observed in recent years possibly due to Ragweed pollen. The study also noted an increase in sales beginning with tree pollen season and also observed a recent increase during August (Rodinkova et.al., 2020). A study conducted in two cities in Sweden saw a high influence of birch pollen levels combined with high air pollution concentration in most of the study locations (Grundström et.al., 2017).

In Serbia, an analysis of the relationship between Birch (Betula) pollen concentration and immunotherapy was carried out for the period from 2015 to 2018. The analysis showed an 80% increase in Birch pollen immunotherapy from 2015 till 2018 and a causal relationship between pollen and immunotherapy. The study also noted that birch pollen was the main allergenic tree species in Serbia (Minić et.al, 2020). Another study done in Nis Serbia noted a high prevalence of asthma symptoms in 6 to 7-year-old children concerning allergic rhinitis (Živković et.al, 2010).

A study that analyzed pollen concentration across cities in various eastern and central European countries including Serbia from 2006 to 2008 noted that Betula made up a significant part of total pollen. The study also mentioned that the mean pollen concentration was more in the Serbian city of Novi Sad compared to other cities in the study which included Timisoara in Romania, Szeged in Hungary, and Ljubljana in Slovenia (Ianovici et.al, 2009). Another study that analyzed seasonal plantago pollen from 2000 to 2004 also noted Novi Sad to have the highest concentration of Plantago pollen among the other cities present in the study which included Timisoara in Romania, Szeged in Hungary, and Ruma in Serbia (Ianovici et.al., 2008).

Methodology

The sales data was obtained from a public dataset that recorded sales of various categories of drugs at a pharmacy in Niš, Serbia (Zdravković et.al., 2020). The dataset recorded 6 years of transactional sales data from 2014 to 2019. The sales of the drugs were grouped into 6 categories using Anatomical Therapeutic Chemical (ATC) Classification System. Out of which R06 denoted Antihistamines for systemic use. This data was recorded hourly and further aggregated into daily, weekly, and monthly levels. We used the daily sales dataset for this study.

The pollen dataset was obtained from Ambee's Gspatial platform for Nis, Serbia (Katti & Srivatsav, 2023). The dataset contains species-wise pollen count as well as overall Tree, Grass, and Weed pollen counts along with their risk levels from 2016-2019. The data were recorded daily.

The trends in pollen and sales of R06 drugs were analyzed at daily, weekly, and monthly levels from 2016 to 2019.

The analysis was done in a jupyter notebook environment using libraries such as pandas, matplotlib, and statsmodels.

The distributions of each of the datasets were looked at. For antihistamine sales, the median number of items sold in a day was 2 with the maximum being 15 (Fig 1). For total pollen count, the median count was 96 pollen grains per cubic meter with the maximum being 978. Both distributions were positively skewed. (Fig 2)

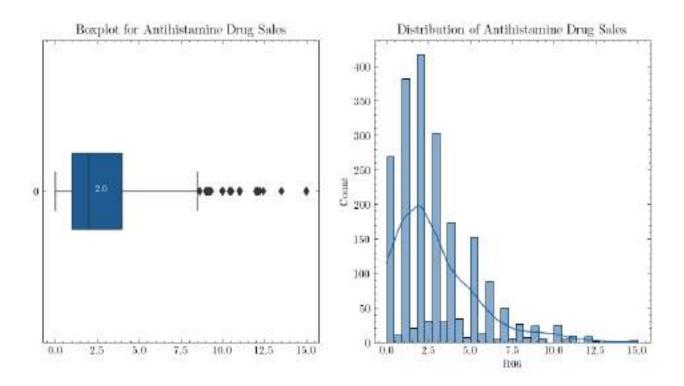


Fig 1: Boxplot and Distribution of Antihistamine Sales

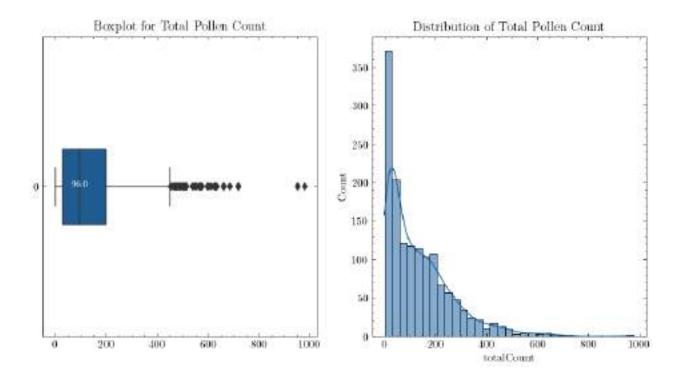


Fig 2: Boxplot and Distribution of Total Pollen Count

To compare values in different scales, both antihistamine sales data and pollen data were normalized and merged using an inner join. The relationship between sales and pollen count was analyzed at a daily level by comparing pollen count and sales on the same day as well as shifting the pollen counts for up to 14 days. For weekly and monthly analysis, the normalized data were aggregated and summed. For weekly and monthly comparisons, data were compared by taking aggregates for the same days as well as taking aggregates after shifting for up to 14 days. Shifting was also done at 10 days intervals and Spearman correlation change was analysed for up to 100 days for weekly comparison. For monthly comparison, data were also compared by shifting it by a week and taking aggregate for up to 4 weeks (28 days).

To analyze the species affecting antihistamine sales, species with less than 20% daily Spearman correlation with R06 were filtered out. Since most of the values for species outside their pollinating season were zero, only non-zero values were taken for this analysis. The trends and correlations were compared at daily, weekly, and monthly levels.

Using statsmodels' OLS package, regression models were developed between total pollen count and antihistamine sales. Using total pollen count as an independent variable, current day and next day's sales were modeled at the daily level, current week and next week's sales were modeled at the weekly level and current month and next month's sales were modeled at the monthly level.

Results

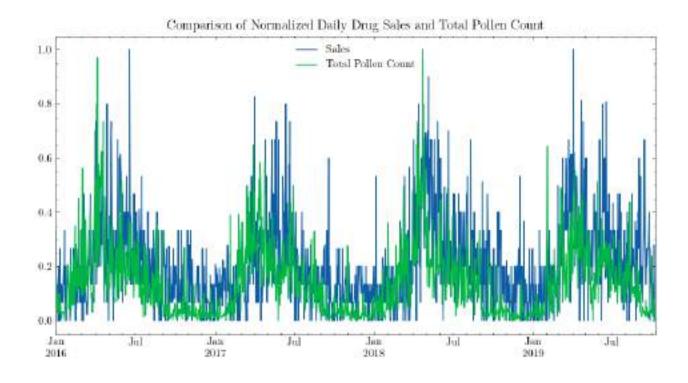


Fig 3: Normalised Trendlines for Daily Total Pollen Count and Sales

For daily data, a similar overall trend was observed. The total pollen count and sales had a moderate spearman R of 41%. Out of Tree, Grass, and Weed, Tree has the highest correlation followed by Grass. Shifting the pollen count for up to 14 days didn't change the correlation too much. At the daily level, both the total pollen count and sales data had higher variance and noise as observed in Fig 3.

	Spearman R with R06
Tree Count	0.271183
Grass Count	0.316217
Weed Count	0.081401
Total Count	0.412610

Table 1: Daily Spearman correlation between R06 and Pollen

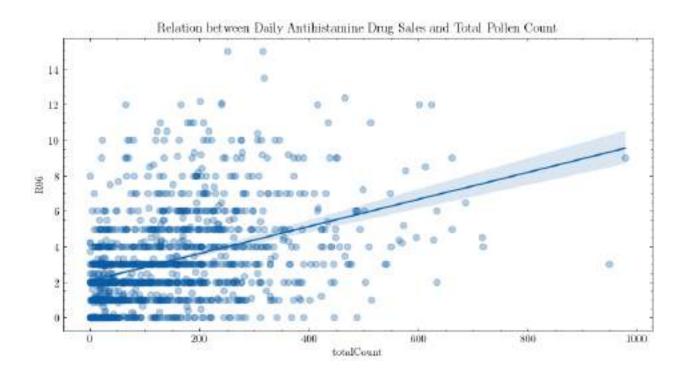


Fig 4: Relationship between daily antihistamine drug sales and total pollen count

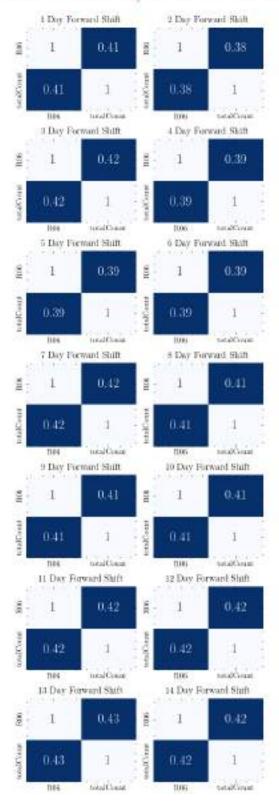


Fig 5: Spearman correlation between shifted total pollen count and sales.

By resampling the data to a weekly sum, A lot of noise was reduced and the data was smoothened out. We saw a significant improvement in correlation from 41% at a daily level to 70% for weekly total sales and pollen count [Table 2].

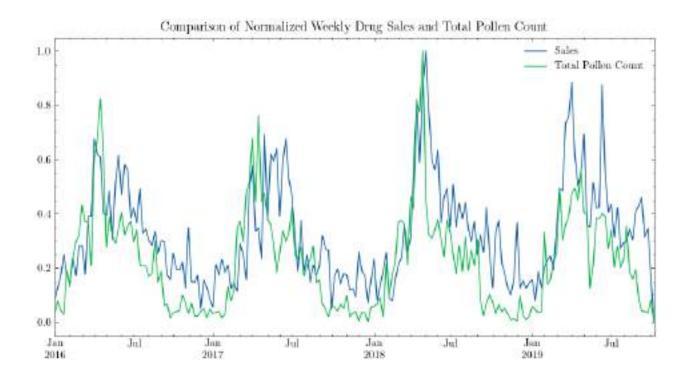


Fig 6: Normalised Trendlines for Weekly Total Pollen Count and Sales

	R06
Tree Count	0.412527
Grass Count	0.499926
Weed Count	0.132132
Total Count	0.691549

Table 2: Weekly Spearman correlation between R06 and Pollen

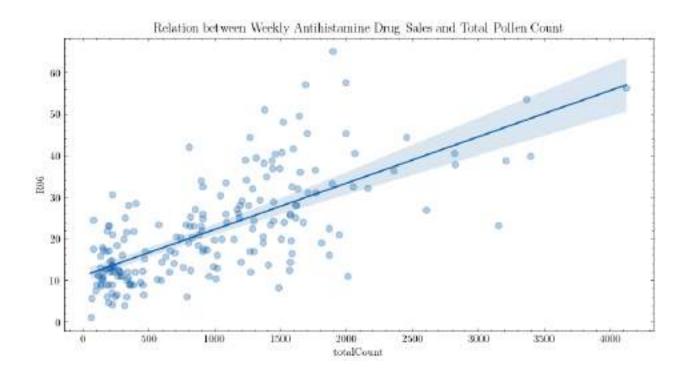


Fig 7: Relationship between weekly antihistamine drug sales and total pollen count

After shifting the pollen data by 8 days and taking a weekly sum, the correlation started to increase. This might suggest that current pollen might affect the sales of antihistamines 8 days later slightly more than the current day's sales. Although the increase in correlation was not very high at 3%.

The total pollen count was shifted for up to 100 days and the weekly sum was taken with a 10-day interval between each shift. After shifting for 10 days and taking a weekly sum, sales had a higher correlation to current pollen but the correlation started to gradually decrease from the 20-day shift and started to rapidly decline from the 60-day shift. Both the sales of Antihistamines and Pollen peak between 10th week to 30th week.

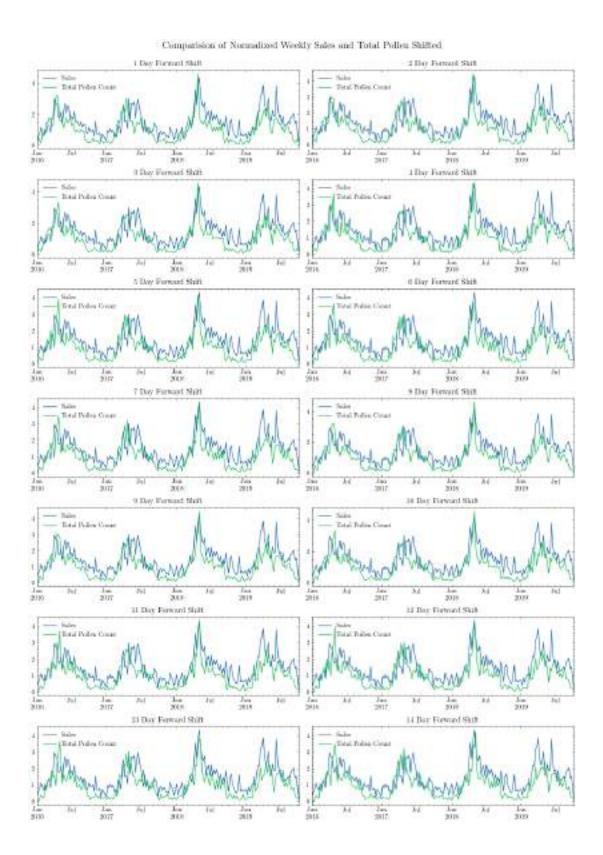


Fig 8: Comparison of normalized weekly sales data and shifted total pollen count

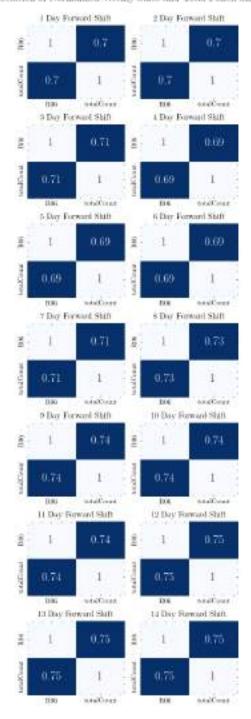


Fig 9: Spearman correlation between shifted total pollen count and sales and summed weekly

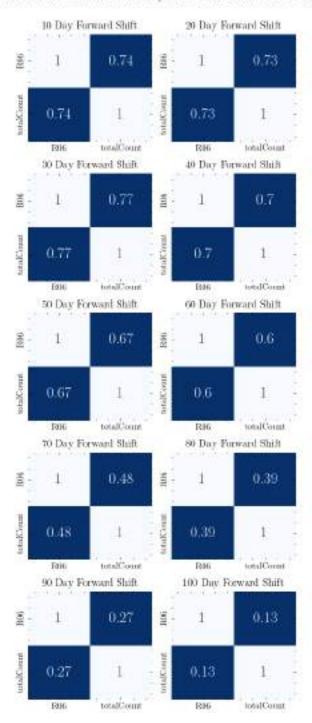


Fig 10: Spearman correlation between shifted total pollen count and sales at 10-day intervals and summed weekly

While pollen sharply declined after the 30th week, the sales of antihistamines declined gradually.

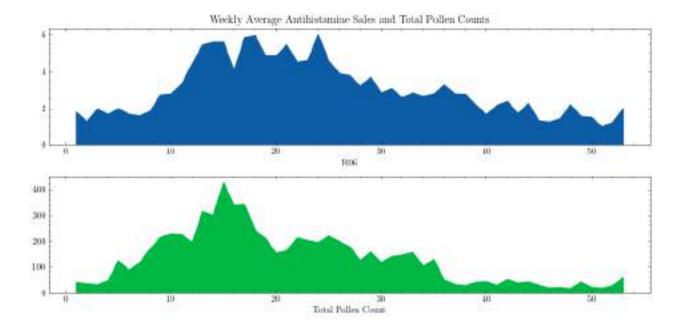


Fig 11: Average antihistamine sales and total pollen counts grouped by week

At the monthly level, the noise is further reduced. We saw further improvement in Spearman R between the monthly sum of sales and the total pollen count at 82%.

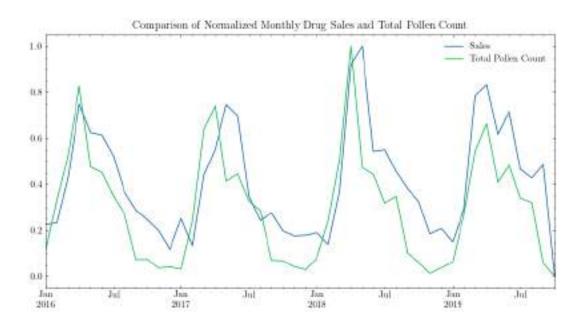


Fig 11: Normalised Trendlines for Monthly Total Pollen Count and Sales

	R06
Tree Count	0.509461
Grass Count	0.521075
Weed Count	0.091234
Total Count	0.817206

Table 3: Monthly Spearman correlation between R06 and Pollen

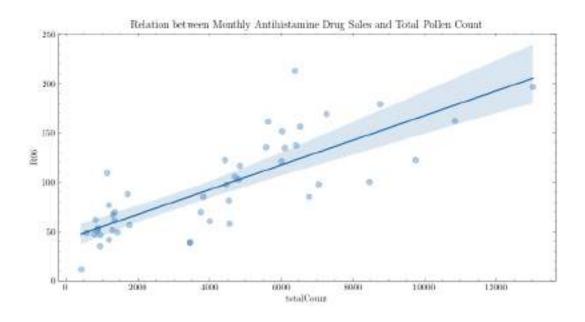


Fig 12: Relationship between monthly antihistamine drug sales and total pollen count

By shifting up to 14 days and taking monthly aggregate the correlation increased up to 88%. With 7-day shifts and taking monthly aggregate, we once again saw that the

Spearman R maxes out at 14 days shift. With approximately 1 month shifted (28 days), the correlation is around 87% which is higher than the current month's correlation of 80%. From the 28-day shift to the 56-day shift, the correlation gradually decreased and after 56 days of shifting, the correlation began to decrease sharply.

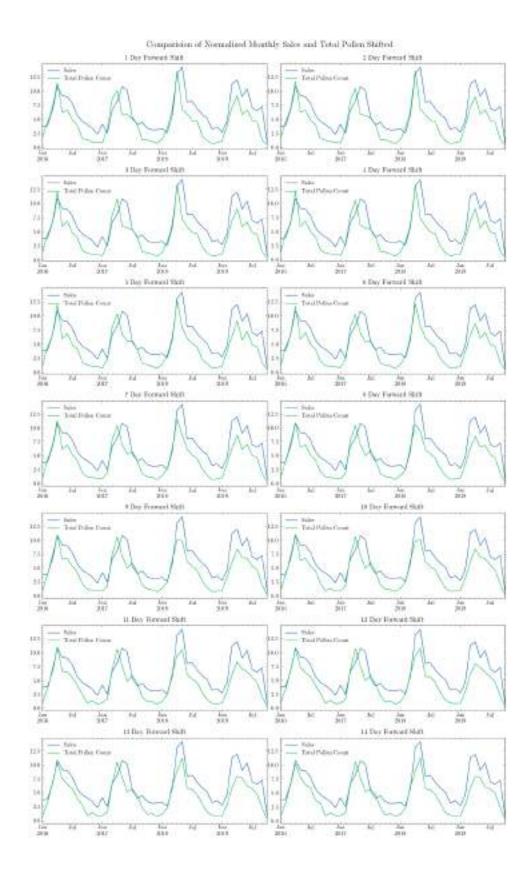


Fig 13: Comparison of normalized monthly sales data and shifted total pollen count

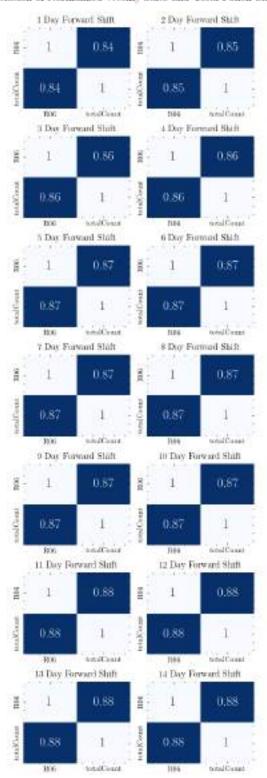


Fig 14: Spearman correlation between shifted total pollen count and sales and summed monthly

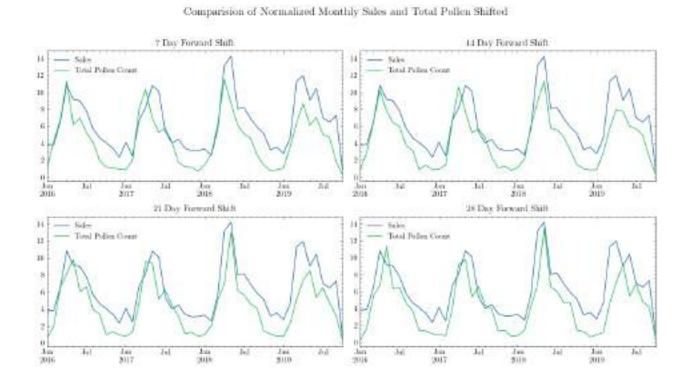


Fig 15: Comparison of normalized monthly sales data and shifted total pollen count with 7 days interval

Correlation of Normalized Weekly Sales and Total Pollen Shifted

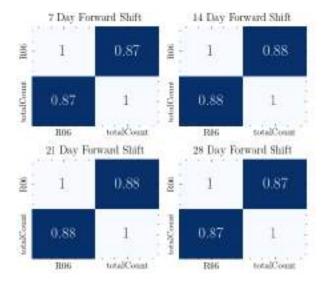


Fig 16: Spearman correlation between shifted total pollen count with 7-day intervals and sales and summed monthly

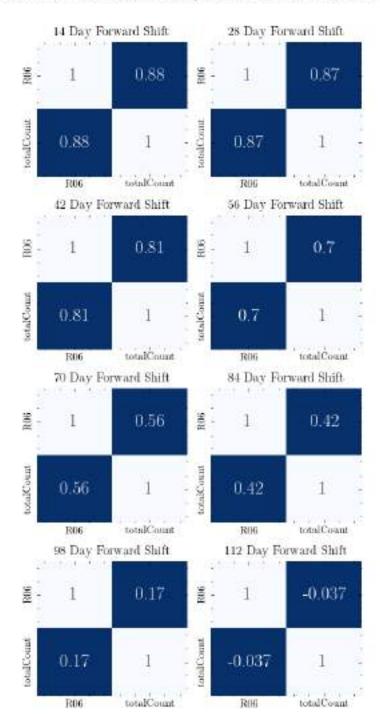


Fig 17: Spearman correlation between shifted total pollen count with 14-day interval and sales and summed monthly

Both Pollen and Sales peaked from March till July with the highest being in April for both of them. After August Pollen sharply reduced while sales decreased gradually.

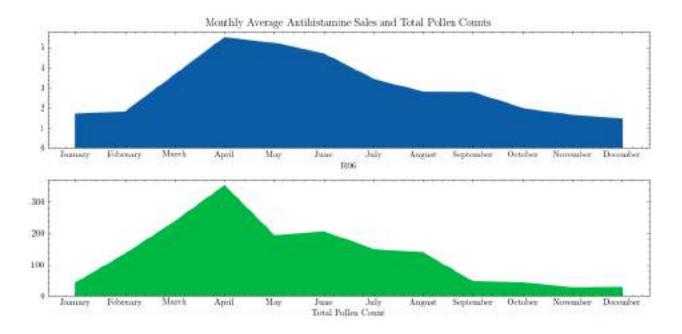


Fig 18: Average antihistamine sales and total pollen counts grouped by month

Oak, Birch and Pine pollen peaked in the months with the highest antihistamine sales. After removing zero values and taking the monthly sum, Oak had a 70% Spearman correlation with sales while Birch had a 60% correlation, Grass (Poaceae) had a 53% correlation and Pine had a 50% correlation with R06.

	R06
Pine	0.500000
Oak	0.709774
Plane	0.423529
Birch	0.595489
Poaceae	0.534788

Table 4: Monthly Spearman correlation between R06 and pollen species

Several regression models were built to model sales using total pollen counts. The model used to fit current-day sales data using current-day pollen count gave an R2 score of 0.157. A mean absolute error was 1.75 which was large for a dataset with average daily sales of 3.14 units. The mean absolute percentage error was 71%. Modeling next-day sales using the current day's pollen count performed similarly with an R2 score of 0.158. The mean absolute error was 1.74 and the Mean absolute percentage error was 72%.

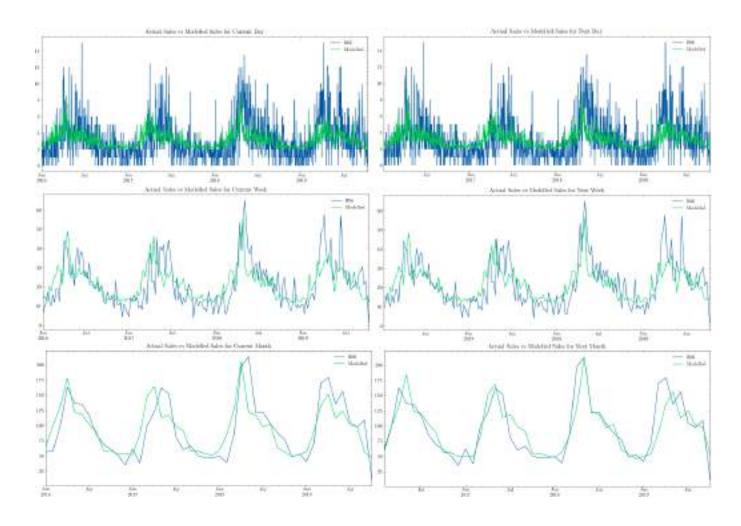


Fig 19: Actual vs Modelled Sales

Weekly models saw a significant improvement over daily models. The model used to fit the current week's sales data using the current week's total pollen count gave an R2 score of 0.48 while the model used to fit the next week's sales gave an R2 score of 0.46. The mean absolute error was 6.63 for the current week's sales and 6.74 for next week's sales with average weekly sales of 21.84 units and 21.92 units respectively. The mean absolute percentage error was at 44% for the current week and 46% for next week's sales.

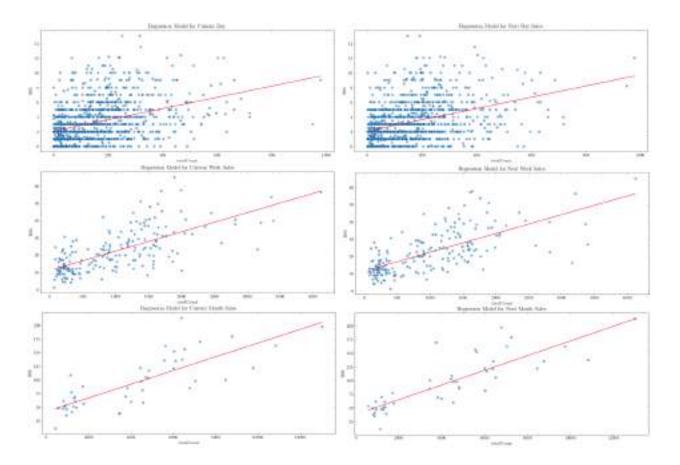


Fig 20: Regression Models

Monthly models performed the best with the model used to fit the current month's sales data using the current month's total pollen count giving an R2 score of 0.64 while the model used to fit the next month's sales gave an R2 score of 0.71. The mean absolute error was 21.88 for the current month's sales and 18.23 for the next week's sales with average weekly sales of 94.00 units and 94.84 units respectively. The mean absolute percentage error was 32% for the current month and 27% for the next month's sales.

Conclusion

Model	R2 Score	MAE (Mean Value)	100 - MAPE
Current Day Sales	0.157	1.749 (3.1403)	28.69%
Current Week Sales	0.479	6.628 (21.840)	55.59%
Current Month Sales	0.636	21.886 (94.006)	67.97%
Next Day Sales	0.158	1.747 (3.143)	28.20%
Next Week Sales	0.460	6.741 (21.920)	54.31%
Next Month Sales	0.712	18.240 (94.840)	73.24%

Table 5: Modelling Summary Statistics

From our observations, we saw that pollen has a significant contribution to antihistamine sales. The best performance was obtained when modeling the next month's sales using the current month's pollen. In general, monthly models performed better than weekly models and weekly models performed better than daily models. Due to increased noise at a daily level for both pollen and sales, modeling sales at a daily temporal resolution did not give optimal results. However, when aggregated to a weekly level, the noise was smoothened out, leading to better models. When we aggregated the data to a monthly level, the noise was smoothened out even further leading to the best results among the different temporal resolutions.

References

- Eli O. Meltzer, Michael S. Blaiss, M. Jennifer Derebery, Todd A. Mahr, Bruce R. Gordon, Ketan K. Sheth, A. Larry Simmons, Mark A. Wingertzahn, John M. Boyle, Burden of allergic rhinitis: Results from the Pediatric Allergies in America survey, Journal of Allergy and Clinical Immunology, Volume 124, Issue 3, Supplement 1, 2009, Pages S43-S70, ISSN 0091-6749, https://doi.org/10.1016/j.jaci.2009.05.013.
- 2. Werchan, M., Werchan, B. & Bergmann, KC. German pollen calendar 4.0 update based on 2011–2016 pollen data. *Allergo J Int* **27**, 69–71 (2018). https://doi.org/10.1007/s40629-018-0055-1
- 3. Lo, F., Bitz, C.M., Battisti, D.S. *et al.* Pollen calendars and maps of allergenic pollen in North America. *Aerobiologia* **35**, 613–633 (2019). https://doi.org/10.1007/s10453-019-09601-2
- 4. Randall, K. L., & Hawkins, C. A. (2018). Antihistamines and allergy. *Australian prescriber*, 41(2), 41–45. https://doi.org/10.18773/austprescr.2018.013
- 5. Comtois, P., Alcazar, P. & Néron, D. Pollen counts statistics and its relevance to precision. *Aerobiologia* **15**, 19–28 (1999). https://doi.org/10.1023/A:1007501017470
- 6. Buters, J.T.M., Antunes, C., Galveias, A. *et al.* Pollen and spore monitoring in the world. *Clin Transl Allergy* **8**, 9 (2018). https://doi.org/10.1186/s13601-018-0197-8
- 7. Polling, M., Li, C., Cao, L. *et al.* Neural networks for increased accuracy of allergenic pollen monitoring. *Sci Rep* **11**, 11357 (2021). https://doi.org/10.1038/s41598-021-90433-x
- 8. Buters, J., Clot, B., Galán, C. *et al.* Automatic detection of airborne pollen: an overview. *Aerobiologia* (2022). https://doi.org/10.1007/s10453-022-09750-x
- 9. Bartková-Ščevková, J. The influence of temperature, relative humidity and rainfall on the occurrence of pollen allergens (*Betula, Poaceae, Ambrosia*

- *artemisiifolia*) in the atmosphere of Bratislava (Slovakia). *Int J Biometeorol* **48**, 1–5 (2003). https://doi.org/10.1007/s00484-003-0166-2
- 10. Peternel, R., Srnec, L., Čulig, J. *et al.* Atmospheric pollen season in Zagreb (Croatia) and its relationship with temperature and precipitation. *Int J Biometeorol* **48**, 186–191 (2004). https://doi.org/10.1007/s00484-004-0202-x
- 11. Zewdie, G. K., Lary, D. J., Levetin, E., & Garuma, G. F. (2019). Applying Deep Neural Networks and Ensemble Machine Learning Methods to Forecast Airborne *Ambrosia* Pollen. *International journal of environmental research and public health*, 16(11), 1992. https://doi.org/10.3390/ijerph16111992
- 12. Csépe, Z., Leelőssy, Á., Mányoki, G. *et al.* The application of a neural network-based ragweed pollen forecast by the Ragweed Pollen Alarm System in the Pannonian biogeographical region. *Aerobiologia* **36**, 131–140 (2020). https://doi.org/10.1007/s10453-019-09615-w
- 13. Notas G, Bariotakis M, Kalogrias V, Andrianaki M, Azariadis K, et al. (2015) Correction: Accurate Prediction of Severe Allergic Reactions by a Small Set of Environmental Parameters (NDVI, Temperature). PLOS ONE 10(4): e0127604.
- 14. Guilbert, A., Simons, K., Hoebeke, L., Packeu, A., Hendrickx, M., De Cremer, K., Buyl, R., Coomans, D., & Van Nieuwenhuyse, A. (2016). Short-Term Effect of Pollen and Spore Exposure on Allergy Morbidity in the Brussels-Capital Region. *EcoHealth*, 13(2), 303–315. https://doi.org/10.1007/s10393-016-1124-x
- 15. Sánchez-Mesa, J. A., Serrano, P., Cariñanos, P., Prieto-Baena, J. C., Moreno, C., Guerra, F., & Galan, C. (2005). Pollen allergy in Cordoba city: frequency of sensitization and relation with antihistamine sales. *Journal of investigational allergology & clinical immunology*, 15(1), 50–56.
- 16. Sheffield, P. E., Weinberger, K. R., Ito, K., Matte, T. D., Mathes, R. W., Robinson, G. S., & Kinney, P. L. (2011). The association of tree pollen concentration peaks

- and allergy medication sales in New York city: 2003-2008. *ISRN allergy*, 2011, 537194. https://doi.org/10.5402/2011/537194
- 17. Ito, K., Weinberger, K.R., Robinson, G.S. *et al.* The associations between daily spring pollen counts, over-the-counter allergy medication sales, and asthma syndrome emergency department visits in New York City, 2002-2012. *Environ Health* **14**, 71 (2015). https://doi.org/10.1186/s12940-015-0057-0
- 18. Rodinkova, Victoria & Yuriev, Sergey & Kryvoviaz, O. & Voronkina, Alyona & Y., Resnik & Yasniuk, Maryna & L.M., Dubuske. (2020). Changes of antihistamine sales patterns in central Ukraine shows the impacts of both global warming and increased ragweed sensitization. 10.1111/all.14508.
- 19. Grundström, M., Dahl, Å., Ou, T. *et al.* The relationship between birch pollen, air pollution and weather types and their effect on antihistamine purchase in two Swedish cities. *Aerobiologia* **33**, 457–471 (2017). https://doi.org/10.1007/s10453-017-9478-2
- 20. Minić, R., Josipović, M., Tomić Spirić, V., Gavrović-Jankulović, M., Perić Popadić, A., Prokopijević, I., Ljubičić, A., Stamenković, D., & Burazer, L. (2020). Impact of Tree Pollen Distribution on Allergic Diseases in Serbia: Evidence of Implementation of Allergen Immunotherapy to *Betula verrucosa*. *Medicina* (*Kaunas*, *Lithuania*), 56(2), 59. https://doi.org/10.3390/medicina56020059
- 21. Živković, Z., Vukašinović, Z., Cerović, S. *et al.* Prevalence of childhood asthma and allergies in Serbia and Montenegro. *World J Pediatr* **6**, 331–336 (2010). https://doi.org/10.1007/s12519-010-0207-y
- 22. IANOVICI, N., JUHASZ, M., KOFOL-SELIGER, A., & SIKOPARIJA, B. (2009). Comparative Analysis of some Vernal Pollen Concentrations in Timisoara (Romania), Szeged (Hungary), Novi Sad (Serbia) and Ljubljana (Slovenia). *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 37(2), 49–56. https://doi.org/10.15835/nbha3723262

- 23. Ianovici, Nicoleta & Eszter, Juhasz & P., Radišič & M., Juhasz & Sikoparija, Branko. (2008). Plantago atmospheric pollinic season in the Danube-Kris-Mures-Tisza Euroregion (2000-2004). Scientific Annals of "Alexandru Ioan Cuza" University of Iasi. (New Series), Section 2. Vegetal Biology. 54. 54-63.
- 24. Zdravković, M., Đorđević, J., Catić-Đorđević, A., Pavlović, S., Ivković, M. Case study: univariate time series analysis and forecasting of pharmaceutical products' sales data at small scale. In: Zdravković, M., Konjović, Z., Trajanović, M. (Eds.) ICIST 2020 Proceedings, pp.1-4, 2020
- 25. Milan Zdravković. (2020). *Pharma sales data* [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DS/466126
- 26. Katti, Pareekshith & Srivatsav, Nithin. (2023). Nis Serbia Pollen. https://doi.org/10.13140/RG.2.2.15181.61922. Data Used for Analysis of the Role of Airborne Pollen in Seasonal Variation of Antihistamine Sales in Serbia