Computational Intelligence

WILEY

Computational Intelligence

SPECIAL ISSUE ARTICLE OPEN ACCESS

Vitamin D Analysis for Sustainable Healthcare in Inner London Borough

Sandra Fernando | Viktor Sowinski-Mydlarz | Subeksha Shrestha | Sunila Maharjan | Duncan Stewart | Dee Bhakta | Gary McLean | Sarah Illingworth

Assistive Technology Research Group, SCDM, London Metropolitan University, London, UK

Correspondence: Sandra Fernando (s.fernando@londonmet.ac.uk) | Sunila Maharjan (sunilamaharjan.2022@gmail.com)

Received: 3 February 2025 | **Revised:** 18 March 2025 | **Accepted:** 25 March 2025

Keywords: data analysis | healthcare analytics | interactive dashboard | predictive modeling | vitamin D prescription patterns

ABSTRACT

Vitamin D is vital for bone health, immune system support, and muscle function. Deficiency in Vitamin D is widespread, with up to 65% of individuals in certain populations, including Black students at London Metropolitan University, UK, being affected. This study focuses on the need for a deeper understanding of Vitamin D prescription patterns, specifically within an inner London borough, using advanced data analytics. Previous analysis, such as ones conducted by OpenPrescribing.net, has investigated NHS prescription data but lacked a focused examination on Vitamin D. Our study introduces a novel computational approach, integrating NHS datasets from 2013 to 2023. We developed a web-hosted dashboard using Python, Flask, Cesium, PowerBI, and libraries such as Pandas, Scikit-learn to provide real-time data visualization and predictive analytics. Our methodology involved API-driven ingestion of large-scale data, focusing on Vitamin D prescriptions in a borough, and mapping this against patient numbers. We used feature manipulation and model training to gain insights into prescription counts, dosages, medication types, and formulations. This interactive platform supports dynamic reporting through PowerBI and Cesium. Our findings reveal significant variations in prescription patterns among GP surgeries influenced by socioeconomic factors. This interdisciplinary project, in future collaboration with local GP federations, United Kingdom, enhances computational health data analysis and aims to inform better prescription practices and healthcare policies, ultimately improving policy practice and public health outcomes.

1 | Introduction

Vitamin D is a fat-soluble vitamin, essential for maintaining healthy bones and teeth, regulating the immune system, and modulating cell growth. Over the years, several studies have investigated the role of Vitamin D in various health effects, including its potential influence on chronic diseases such as cardiovascular disease, diabetes, and cancer [1] Understanding the methodologies and statistical approaches used in these studies is crucial for evaluating their validity and applicability to broader populations.

The study on Vitamin D prescription patterns in this study used a combination of descriptive and predictive statistical methods to analyze and interpret the data. Descriptive statistics provided a summary of the prescription data, including total counts, dosages, medication types, and formulations. We performed feature manipulation to obtain information on drug formulations, dosages, patient ratios, medicine names, time series data, prediction data, and items per 1000 patients. We also computed totals for prescriptions, counts, filters, time durations, surgeries, and formulations.

This analysis allowed the identification of patterns and variations in Vitamin D prescriptions across different GP surgeries, determining a baseline understanding of Vitamin D prescribing practices. This foundational data will support further research to the antioxidative effects of Vitamin D on telomere length (TL) in an in vitro cell model. [2] Validate the use of Vitamin D rapid point-of-care tests using an Igloo device [3]; Gather stakeholder perspectives on the implications of our findings.

However, this article primarily discusses the development and implementation of the Vitamin D dashboard and data pipeline, emphasizing its strengths and challenges from a Data Science perspective.

2 | Literature Review

2.1 | Key Researchers and Their Studies

Prior attempts at creating dashboards using open-source health datasets, such as OpenPrescribing.net, have shown substantial utility in visualizing and analyzing NHS prescription data. OpenPrescribing.net (Figure 1), developed by the Bennett Institute for Applied Data Science at the University of Oxford, provides an interactive tool helping users explore NHS GP prescribing behaviors across England. This platform allows detailed analysis of over 700 million rows of data, offering an understanding of prescription patterns, costs, and safety measures. Its strengths are its accessibility, real-time updates, and ability to identify outliers and trends at the practice, regional, and national levels. While Oxford's dashboard excels in providing a wide scope of data for general trends and patterns across the United Kingdom [4], our focus on detailed, localized data allows for more precise and actionable insights for local healthcare providers. This localized approach enables targeted interventions and more efficient resource allocation within the community. Some of the enhancements to OpenPrescribing.net include advanced analysis of the BNF_DESCRIPTION column, improved visualizations, and a comprehensive Quick Summary of important data over an extended period. Additionally, the platform now features 3D visualizations, forecasting capabilities, and exportable reports, all of which are explained in Section 6.

ePACT2, run by the NHS Business Services Authority, is a formidable tool that offers an in-depth understanding of how NHS prescriptions are administered. It is great for organizations that require detailed financial analysis, tracking trends, and comparing practices across different regions. One significant advantage is its ability to highlight uncommon prescribing patterns, which can assist in targeting specific issues. However, the downside is that ePACT2 mainly focuses on financial data, so it might not be as useful for more clinical or qualitative insights. Also, its complexity might be overwhelming for those who are not used to handling such detailed data [5].

OpenSAFELY was created during the COVID-19 pandemic to analyze large-scale health data, including NHS electronic health records. It is, particularly, effective in handling big datasets securely, making it a good fit for researchers and policymakers concerned about public health issues. The platform's focus on transparency and open research methods adds to its credibility. However, because it covers a wide range of health data, it does not focus enough on the specifics of prescription data. Additionally, navigating and interpreting the data may involve a high level of technical expertise [6].

PrescQIPP is a not-for-profit service designed to help NHS commissioners improve prescription practices. It includes practical tools like reports, bulletins, and case studies that are particularly useful for making prescriptions safer, more effective, and more

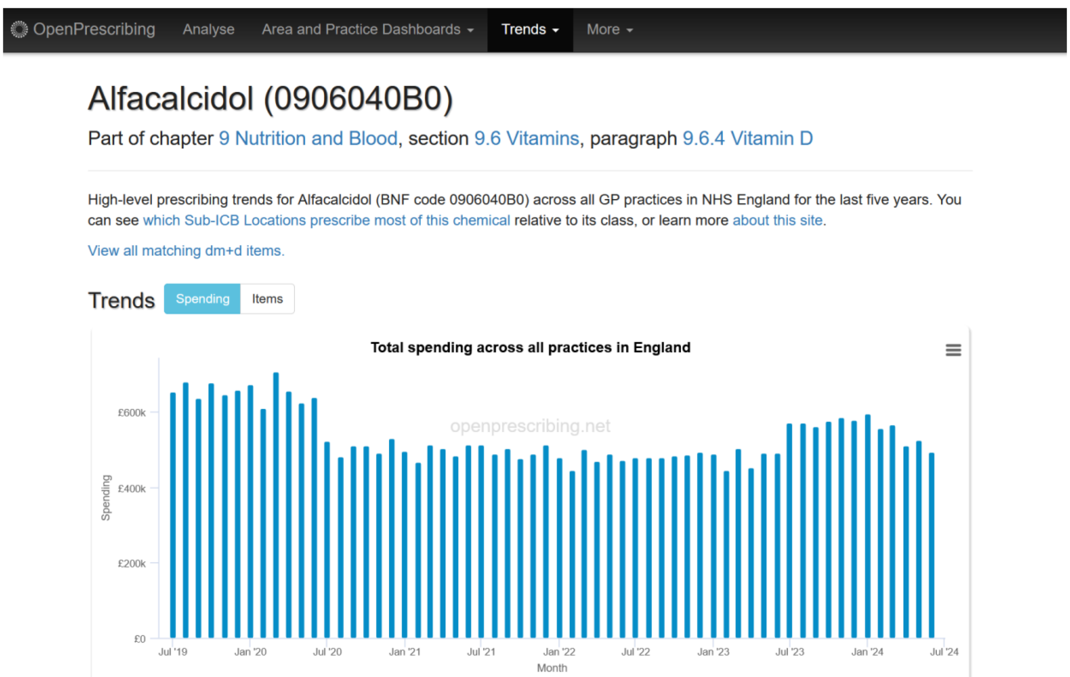

FIGURE 1 | OpenPrescribing dashboard—Alfacalcidol.

cost-efficient. It also provides effective educational resources and support networks, making it a good option for those in charge of overseeing prescription practices. On the downside, PrescQIPP might not offer as many advanced analytical tools or detailed data visualizations as some of the other platforms. This could be a limitation for users needing more comprehensive data analysis [7].

Nonetheless, the challenges linked to such dashboards include ensuring data accuracy, managing incomplete datasets, and maintaining user-friendly interfaces that support the complexity of health data. Furthermore, interpreting the data needs careful consideration of contextual factors, such as patient demographics and practice-specific circumstances, to avoid misinterpretations [8].

HealthData.gov provides a wide range of health-related datasets and visualization tools, allowing users to create custom charts and maps. These platforms enhance data transparency and support evidence-based decision making in healthcare, though they require continuous updates and maintenance to remain relevant and effective [9].

The Institute for Health Metrics and Evaluation (IHME) provides a suite of interactive tools on its website, such as the Global Burden of Disease (GBD) Compare. It allows users to analyze health trends over time and by region. These tools are designed to handle large and complex datasets, giving an overview of global health patterns and enabling policymakers to make informed decisions [10].

Dr. Michael F. Holick is a prominent figure in Vitamin D research, acclaimed for his extensive work on its benefits for bone health and chronic diseases, though some critics signal that he may overemphasize deficiency rates and risks of toxicity [11]. Dr. Bruce W. Hollis's research on Vitamin D metabolism and its effects on pregnancy and infant health is well-regarded, despite criticisms about small sample sizes and short follow-up periods affecting generalisability [12]. Dr. Adrian R. Martineau's studies on Vitamin D supplements and respiratory infections are lauded for their design, but meta-analyses yield mixed results due to varying dosages and baseline Vitamin D levels [13]. Dr. JoAnn E. Manson's large-scale studies, such as the VITAL trial, have significantly contributed to understanding Vitamin D's role in preventing chronic diseases, although interpretation of results is challenged by participant inconsistency in baseline levels and adherence [14]. Dr. Heike A. Bischoff-Ferrari has shown Vitamin D's positive impact on musculoskeletal health in elderly adults yet faces critiques regarding participant selection biases and dosage irregularities [15].

Visualization tools like QlikView, D3.js, and Plotly are commonly used for healthcare data visualization. QlikView offers a drag-and-drop interface and associative search capabilities, making data exploration intuitive. D3.js allows high customization for creating dynamic visualizations, while Plotly supports multiple programming languages and a wide range of interactive features. These tools enhance the ability to visualize complex health data, though each has its own strengths and limitations depending on the specific requirements of the healthcare organization [10].

Our study builds on these capabilities by developing a robust Vitamin D prescription dashboard that addresses previous limitations. We focus on enhancing data analytics capabilities through advanced computational techniques and providing valuable insights into prescription practices and their socioeconomic associates, ultimately aiming to inform better healthcare policies and improve public health and patient outcomes.

3 | Methodology

This research project intends to investigate patterns of Vitamin D prescriptions in the borough, focusing on various characteristics such as dosage variations, the integration of point-of-care tests, and the antioxidative effects on TL. Below is the detailed methodology, along with the rationale behind the chosen frameworks, languages, storage solutions, and datasets.

Numerous portals and web interfaces have been employed to present health data. They apply diverse statistical and visualization techniques to enhance data accessibility and interpretation [16]. A notable example is OpenPrescribing.net, which offers interactive tools for exploring NHS GP prescribing data across England. This platform features real-time data updates, comprehensive dashboards, and customizable analyses, enabling users to identify trends, outliers, and potential cost-saving opportunities. The primary strengths of such interfaces are their accessibility, user-friendliness, and ability to manage large datasets effectively. They also face many challenges, including data accuracy, incomplete datasets, and designing intuitive interfaces that accommodate a wide range of users.

A critical component of this study was data preprocessing, which involved cleaning and preparing the data to ensure its accuracy and relevance. This process included feature manipulation to extract key variables, such as drug formulations, medicine names, medication types, dosages, and patient ratios. By refining these variables, the study was able to conduct more reliable analyses and generate more accurate predictions. This meticulous approach to data preparation not only enhances the reliability of the results but also supports the development of more sophisticated predictive models, ultimately contributing to better-informed healthcare policies and practices.

In this research, predictive analytics, including machine learning techniques and regression analysis, was employed to forecast future trends based on historical data. These models are invaluable for planning and optimizing prescription practices, as they provide insights into potential future patterns and help anticipate changes in Vitamin D prescription needs. The use of predictive analytics enables healthcare providers to make informed decisions, ensuring that they can proactively address anticipated demands and improve patient outcomes.

3.1 | Critique of the Methods

Creating a web interface to visualize health data, particularly on vitamin D prescription patterns, involved several basic steps to ensure comprehensive and user-friendly access to data insights. This project, as demonstrated by the analysis of vitamin D prescriptions, integrates data from various sources including NHS API calls, English Prescribing Data, and geospatial data from Open Street Maps, open-source tools such as Python, Flask, libraries such as pandas, NumPy, scikit-learn, and visualization tools such as PowerBI and Cesium to build a dynamic, interactive dashboard. This platform offers real-time data visualizations on prescription patterns, chemical types, formulations, dosages, and predictive analytics.

API calls proved crucial in our research by enabling real-time data access, which was necessary for analyzing up-to-date prescription trends and patterns. The automation provided by APIs substantially reduced the time and effort needed to gather large datasets, allowing us to focus more on the analysis and interpretation of the data. Moreover, APIs offer scalability and flexibility, accommodating large volumes of data and allowing us to filter and request specific information relevant to our study. This capability was particularly beneficial in handling the extensive and dynamic nature of the NHS datasets we utilized, warranting that our analysis was both comprehensive and up to date.

However, the use of APIs also presented several challenges that required careful management. Ensuring data accuracy and consistency was a major concern, as APIs can sometimes return incomplete and/or outdated information, which could compromise the reliability of our findings. We also had to navigate rate limits imposed by the APIs, which restrict the number of calls we make within a certain timeframe, occasionally leading to delays in data retrieval. Additionally, integrating data from multiple sources required significant preprocessing to ensure compatibility and consistency across datasets.

Flask proved to be a highly effective tool in our research, providing a lightweight and flexible framework for developing our web-based dashboard. One of the significant strengths of Flask is its simplicity, which allowed us to quickly set up a robust web application to visualize and interact with our Vitamin D prescription data. Flask's modular design and extensive documentation made it easy to integrate various Python libraries, such as Pandas, NumPy, and matplotlib, for data processing and visualization. This flexibility enabled us to create a customized dashboard tailored to the specific needs of our project, facilitating real-time data access and dynamic reporting.

However, we also encountered several challenges while using Flask. Being a micro-framework, Flask is minimalistic by design, which, while beneficial for simplicity, can pose limitations when scaling the application or incorporating more complex features. Integrating Flask with other tools, such as PowerBI, for data visualization posed challenges, including managing browser restrictions and login requirements, which complicated external access. While Flask served our needs well for this project, scaling it for larger, more complex applications would require significant enhancements or potentially transitioning to a more feature-rich framework.

The use of Cesium for 3D geospatial content enables detailed street map views, enhancing the visualization of prescription trends and healthcare disparities across different regions. Cesium was a powerful tool in our project, particularly, for its ability to render complex geospatial data in a highly interactive and visually compelling way. One of its key strengths is the capacity to handle and visualize large-scale 3D geospatial content efficiently, which allowed us to map Vitamin D prescription patterns across the borough with remarkable detail. Cesium's ability to integrate seamlessly with various data sources provided a dynamic platform for exploring geographic trends.

However, working with Cesium also introduced several challenges, particularly, related to the complexity of setting up and managing the 3D visualizations. The steep learning curve associated with Cesium required significant time and effort to master, especially when integrating it with other tools and datasets. Performance was another concern, as rendering large geospatial datasets in real time could strain system resources, leading to potential lag or delays in visualization, particularly, on lower-end devices.

ARIMA and SARIMA models were used as predictive analysis in our research. These are statistical methods used for time series forecasting. The SARIMA model has a seasonal component as well, so that it handles the seasonality factor in forecasting. However, further research is required for predictive analysis.

Additionally, the framework supports periodic reporting and integration with human cell model validations, thereby facilitating comprehensive analysis and application in real-world healthcare settings. This research not only aids in understanding prescribing habits but also promotes the use of innovative point-of-care tests, aiming to improve healthcare efficiency and patient outcomes.

4 | Implementation

4.1 | Analysis Scope

The technical summary and categorization aspects of Vitamin D prescription data analysis focus on the detailed examination of various formulations and dosage types. The study categorizes Vitamin D formulations into capsules, injections, liquids, and other types, and further breaks down dosages into units such as micrograms, milligrams, milliliters, and standardized units. These dosage ranges are specifically delineated, for instance, from 0 to 1000 units, 0 to 0.25 μg, and 0 to 2000 mL.

The analysis includes detailed data on the types of Vitamin D drugs prescribed within surgeries, along with their specific dosage measurements and formulations. It examines the number of times each item is prescribed by GP surgeries, the total quantity of these prescriptions over different years and months, and the associated costs, including both net ingredient costs and actual costs. Additionally, the study projects the costs, total quantities, formulations, and dosage trends for the next 5 years, providing valuable insights for future healthcare planning and policy development.

The study also addresses several technical challenges encountered during the data analysis process. For example, it describes the efforts to generalize patient counts using digital NHS data and resolves compatibility issues with PySpark when used in conjunction with Flask. Furthermore, the integration of PowerBI

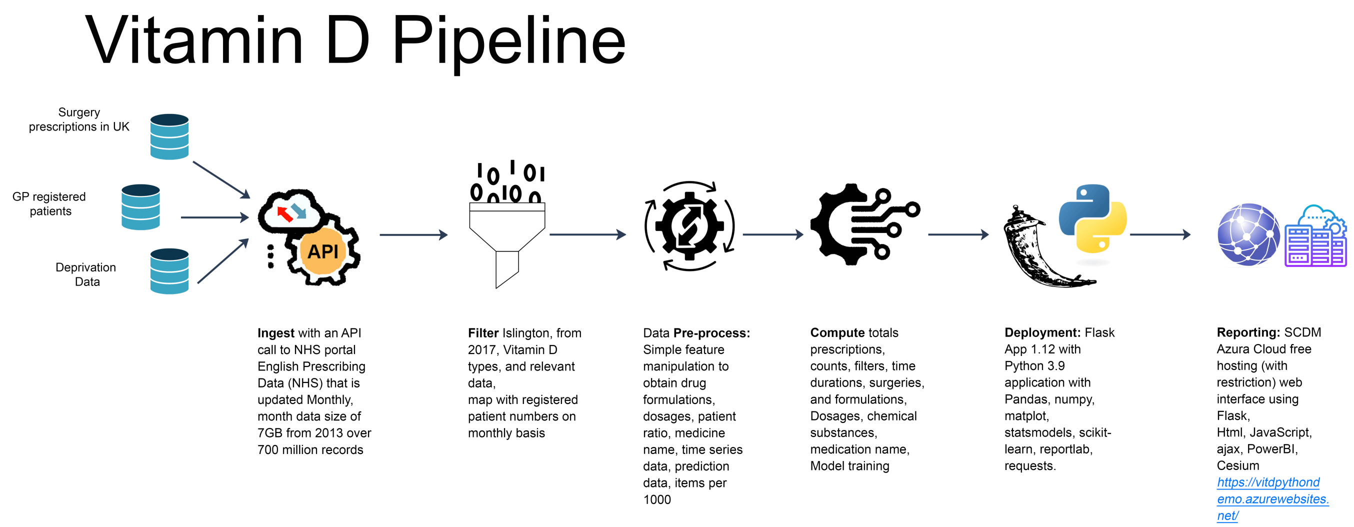

FIGURE 2 | Vitamin D pipeline.

charts for external users presented browser restrictions and login requirements, which were also managed effectively. The technical framework of the study relies on open-source tools and libraries, including Python, Flask, Pandas, and the requests library for making API calls and feature manipulation.

Figure 2 outlines the "Vitamin D Pipeline," which details the comprehensive process used to collect, process, analyze, and visualize Vitamin D prescription data in the borough. The figure illustrates the journey of data from initial ingestion via an API call to the NHS portal, capturing data on surgeries, GP-registered patients, and deprivation data. It highlights how English Prescribing Data, updated monthly, is filtered to focus on Vitamin D prescriptions relevant to the borough from 2017 onward. The pipeline includes data preprocessing steps, such as feature manipulation to extract key variables like drug formulations/dosages, etc. and using model training for predictive analysis. The processed data is then developed in a Flask application and deployed in Azure Cloud, integrated with tools like PowerBI and Cesium for reporting and visualization, providing a dynamic, interactive platform for stakeholders to engage with the findings. The pipeline is a crucial part of the project, ensuring that the data is effectively transformed into actionable insights.

4.2 | Data Ingestion

The data for this project comes from the NHS portal API, and GP-registered patient data. Using the NHS API safeguards that the data is up-to-date and comprehensive, capturing the latest prescription information. GP registration data is crucial for ensuring that the analysis is relevant to the local patient populations in the borough. To gather this data, we used the Python "requests" library for making API calls.

The Python "requests" library is a powerful tool for facilitating HTTP requests, which are how data is exchanged over the internet. Basically, it allows a program to send a request to a web server and retrieve the response. This library abstracts the complexities involved in making these requests, providing a straightforward and user-friendly interface for tasks such as accessing web pages, retrieving data from APIs, and submitting information to web services.

4.3 | Data Storage

Data are stored using a combination of Azure Cloud and local secure servers. Azure Cloud is selected due to its scalability and robust security features, which are crucial for handling sensitive health data. Additionally, local servers offer enhanced security and control over data access, complementing the cloud storage solution to ensure data integrity and security.

Azure Cloud and local secure servers together provide a robust data management system that accommodates large datasets, ensures high-speed data processing, and handles diverse data types. Azure's scalable resources and pay-as-you-go model offer flexibility and cost efficiency, while local servers enhance processing power and allow for simple expansion. The setup supports both real-time data transfer and secure storage, with Azure offering advanced security features like encryption and regulatory compliance, and local servers providing enhanced physical and network security through firewalls and regular audits.

4.4 | Data Processing

The project utilizes Python libraries such as Pandas, NumPy, and so forth, for data processing, while acknowledging the suitability of Apache Spark for larger datasets. Initially, Pandas and NumPy were chosen for their efficiency in data manipulation and numerical operations on smaller datasets, offering the flexibility needed for detailed analysis and feature extraction. However, as the dataset grows, Apache Spark remains a viable option due to its ability to handle large-scale data efficiently, making it an excellent choice for big data analytics required in this study.

4.5 | Data Analysis

For data analysis, we employ time series analysis, statistical modeling, and machine learning techniques to identify trends, patterns, and correlations within the prescription data, offering deeper insights into prescribing behaviors and their potential health impacts.

1. Time Series Analysis: This technique analyzes data points over time to detect trends and seasonal variations, crucial for understanding changes in prescription patterns. It can be complex due to the need for handling trends, seasonality, and noise.
2. Statistical Modeling: Models like linear regression quantify relationships between variables, aiding in hypothesis testing and prediction. They provide a solid framework but rely on assumptions that may not always hold true.
3. Machine Learning Techniques: Algorithms from Scikit-learn build predictive models that learn from data and uncover hidden patterns. These models are robust but can be prone to overfitting and require large datasets and parameter tuning.

Tools Used

- Statsmodels: Ideal for statistical modeling and time series analysis, offering comprehensive tools for estimation and testing, though it can be less user-friendly.
- Scikit-learn: A versatile library for building and evaluating machine learning models, known for its simplicity and efficiency.
- Matplotlib: Used for creating static, animated, and interactive plots, highly customizable but with a steep learning curve for advanced visualizations.

4.6 | Data Visualization

Data visualization is performed using PowerBI and Cesium for geospatial data. PowerBI is chosen for its interactive and comprehensive data visualization capabilities, making it easier to present findings to stakeholders in an understandable and actionable manner. Cesium is used for advanced geospatial visualizations, providing a detailed view of prescription patterns across different areas.

4.7 | Reporting

The reporting component of the project is managed using a Flask web application and ReportLab for generating PDFs and CSVs, chosen for their specific strengths and suitability for the project's needs. Flask, selected for its simplicity and efficiency, enables real-time data access and interaction through a lightweight and flexible framework that is easy to set up and well-supported by extensive documentation and a vibrant community. This choice has proven effective for the rapid development and deployment of web applications. ReportLab was chosen for its powerful capabilities in creating detailed PDF reports, ensuring all relevant information is documented and easily shareable with stakeholders. It excels in generating professional-quality documents with complex layouts and dynamic data integration. However, there are considerations to be aware of Flask may require additional extensions for advanced features and might face scalability issues with very large applications, while ReportLab has a steep learning curve for designing intricate layouts and can be resource-intensive when generating large PDFs. Despite these challenges, the combination of Flask and ReportLab offers a robust, flexible, and effective solution for comprehensive project reporting.

4.8 | Sample Size Verification

The verification of the sample size using statistical power analysis involved several meticulous steps to ensure the study's reliability and validity. The process began by collecting and preprocessing a large dataset obtained from NHS API calls, specifically targeting Vitamin D prescriptions in the area. This data, spanning from 2013 to 2023, was filtered and mapped to registered patient numbers on a monthly basis, providing a detailed understanding of the local population. The sample size was further refined by cross-referencing this NHS data with local GP records, ensuring that it accurately reflected the demographic and health profiles of the population. The generalization of patient counts was resolved by linking the data to the number of patients registered at GP surgeries on the first day of each month. After ensuring the accuracy and reliability of the data, statistical power analysis was conducted to confirm that the sample size was adequate to detect significant effects. This process included iterative checks and adjustments to address any discrepancies, ultimately ensuring a statistically robust and representative sample for the study.

4.9 | Dashboard Creation Steps

The creation of the dashboard involves several steps, starting with requirement analysis to identify key metrics and visualizations needed by stakeholders. Data collection is then performed using the NHS API, integrating deprivation data and GP registration data. We initially tested our data processing with Apache Spark but ultimately chose Python due to its capability to efficiently handle datasets up to 11 GB, which was sufficient for our project needs. Python's libraries, such as Pandas and NumPy, provided the flexibility required for our analysis without the additional complexity of Spark. Additionally, the Dask framework presents a strong alternative to Spark for parallel processing and handling larger datasets within the Python ecosystem, offering a more straightforward integration with our existing workflow [17].

Following data processing, time series analysis, and statistical modeling are conducted to compute prescription counts and filter data by time durations, surgeries, and formulations. The dashboard is then developed using PowerBI for interactive visualizations, with geospatial maps created using Cesium.

Deployment involves hosting the dashboard on a Flask web application on Azure Cloud. Reporting and integration are managed using ReportLab for detailed PDF reports, which are integrated into the dashboard. User testing sessions are conducted with stakeholders to collect feedback and refine the dashboard. Figure 3 explains our project flow.

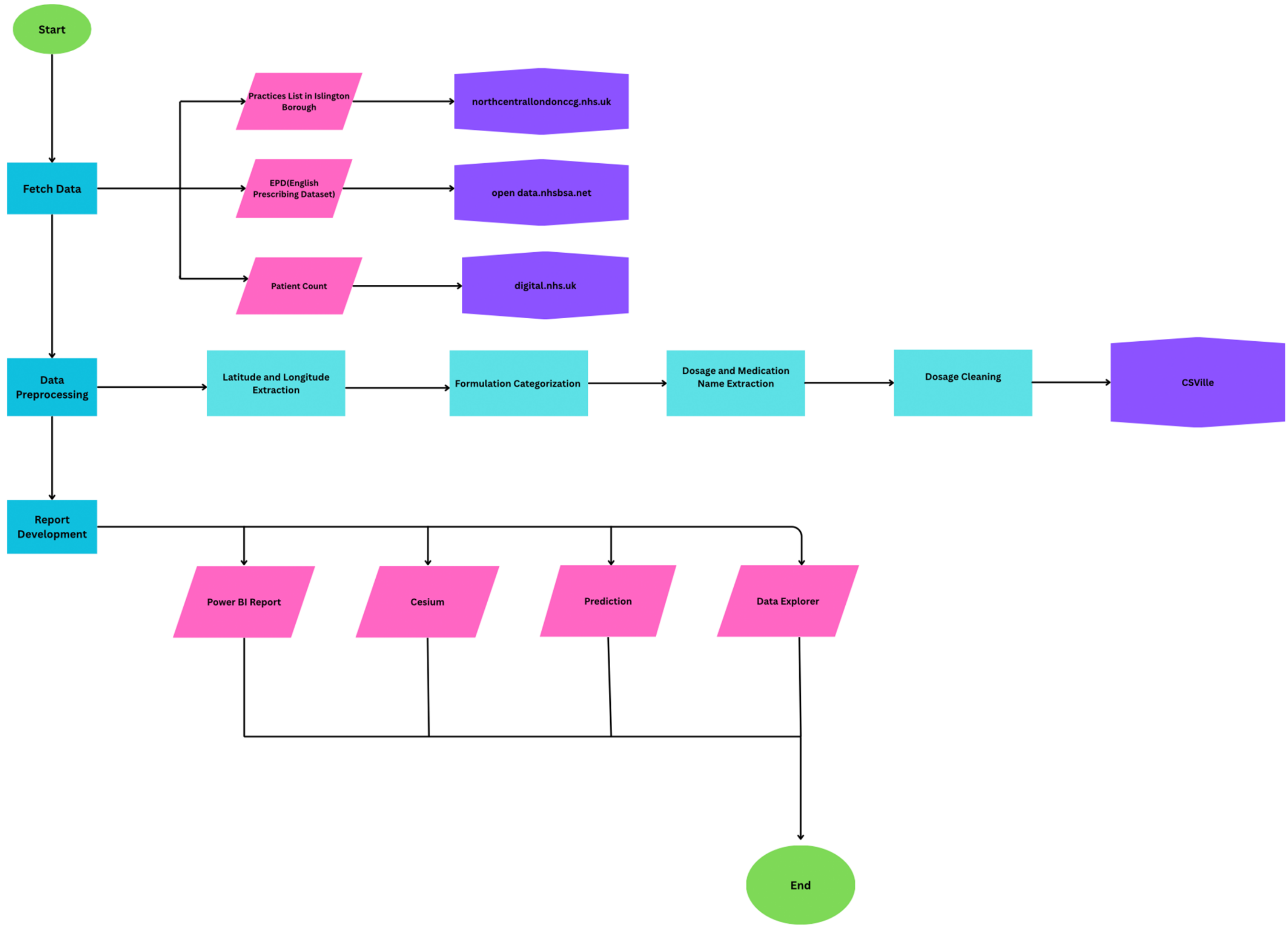

FIGURE 3 | Project flowchart.

4.10 | Key Decisions and Rationale

The choice of frameworks and languages was determined by the need for efficiency, scalability, and robust data-handling capabilities. Python was chosen, with Pandas and NumPy providing powerful data manipulation capabilities suitable for the borough-specific dataset. However, had the project encompassed data for the entire United Kingdom, Apache Spark would be a more appropriate choice due to its ability to process large datasets effectively. PowerBI was chosen for its advanced data visualization features, and Flask was selected for its ease of deployment for web applications. Azure Cloud was chosen for its scalability and security features.

NHS Prescription Data provides up-to-date information on prescribing patterns. The final fields selected for the project were those that provided comprehensive information on prescription patterns, including the BNF description (the name given to the specific type, strength, and formulation of a drug), item count, measurements, and the costs associated with these prescriptions. These columns were chosen because they offer detailed insights into what types of Vitamin D were prescribed, in what quantities, and at what cost, which are critical for understanding prescribing behaviors and financial impacts.

Additionally, the column containing the *BNF description* was text-mined further. The BNF description column in NHS prescription data refers to the detailed textual representation of a medication's name and formulation based on the British National Formulary (BNF) classification. This column was chosen because it includes detailed information about the drug formulations, dosages, and types, which are essential for analyzing the variations in prescriptions and understanding the prescribing patterns in depth. By text-mining this column, we could categorize the formulations and dosages more accurately, thereby enhancing the quality of our analysis and the insights derived from it.

GP Registration Data ensures the relevance and accuracy of the analysis to local patient populations. This project leverages a robust and interdisciplinary approach, combining big data analytics, advanced visualization techniques, and real-world healthcare applications to understand and improve Vitamin D prescription practices. The use of open-source tools and platforms ensures cost-effectiveness and accessibility, fostering collaboration and innovation in the research community.

5 | Results

The dashboard for Vitamin D prescription patterns features several interactive charts and graphs, providing comprehensive

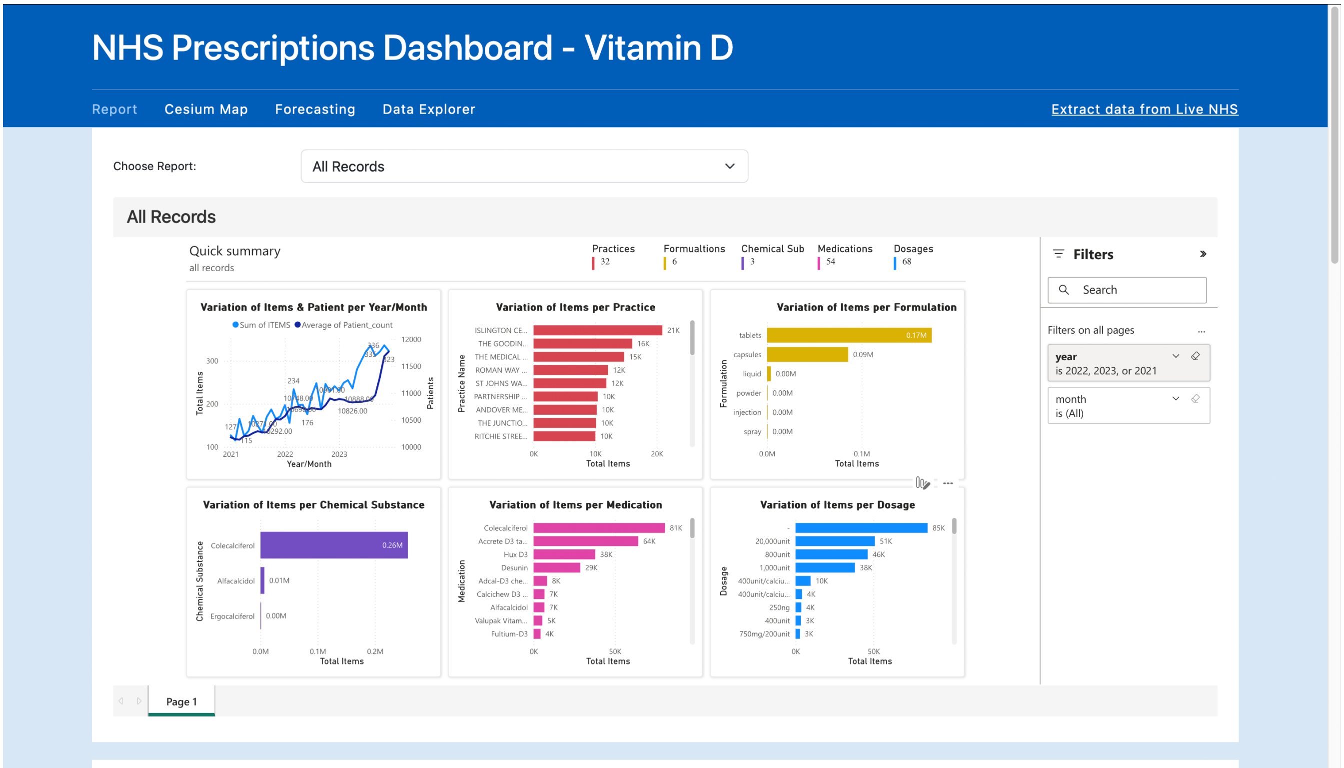

FIGURE 4 | NHS prescriptions dashboard—Vitamin D (all records).

insights into the data. Key elements include time series graphs showing trends in prescription volumes over time, bar charts categorizing prescriptions by type, formulations, and dosage, and geographic maps displaying distribution patterns across different regions in the borough. Predictive graphs forecast future prescription trends based on historical data, offering valuable projections for healthcare planning.

Figure 4 shows an interactive tool designed to visualize various aspects of Vitamin D prescriptions, features a navigation bar at the top with options such as "Report," "Cesium Map," "Forecasting," and "Data Explorer," along with a link to "extract live data from the NHS." The dashboard allows users to select different reports through a dropdown menu, currently set to "All Records," to filter and view data accordingly. A quick summary section provides an overview of key metrics, including trends in the number of items prescribed per patient over time, variations in prescriptions across GP practices, and differences in formulations, chemical substances, medications, and dosage levels. These insights are presented through various visualizations like line graphs and bar charts. Additionally, a filter panel on the right enables users to refine the data by selecting specific years, months, and other criteria. Apart from this, there are eight other reports with visualizations like, line graphs, bar charts, stacked bar charts, donut charts, and maps with individual filter options. This dashboard offers a comprehensive analysis of Vitamin D prescription patterns, facilitating the identification of trends and variations in prescribing behaviors.

5.1 | Key Charts and Findings

Time Series Analysis—these graphs illustrate how Vitamin D prescriptions have fluctuated over the years. They highlight

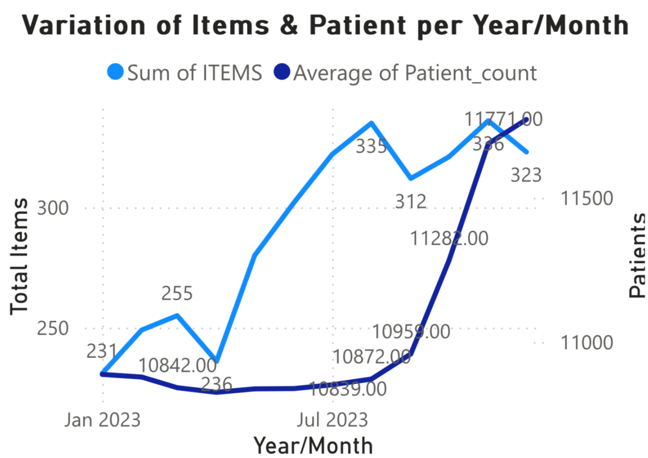

FIGURE 5 | Time series analysis.

seasonal variations and potential correlations with public health initiatives or changes in medical guidelines. For example, in Figure 5, we observed an increase in prescriptions during the winter months, likely due to decreased sunlight exposure and increased awareness of Vitamin D deficiency during these periods.

Bar Charts—these provide a detailed breakdown of the different types of Vitamin D formulations prescribed, such as tablets, capsules, injections, and liquids, along with their respective dosages. In Figure 6, the data show that tablets are the most prescribed form, followed by capsules, demonstrating a preference for oral administration methods in the area.

Geographic Maps—using data from Open Street Maps, these visualizations map out prescription patterns across the borough, showing concentrations of high and low prescription rates and correlating them with socioeconomic data. In Figure 7, areas with higher deprivation levels have higher prescription rates, suggesting a link between socioeconomic status and Vitamin D deficiency.

Predictive Graphs—these projections use historical data to forecast future trends in Vitamin D prescriptions. They help to anticipate demand and inform resource allocation in healthcare services. In Figure 8, the predictive models ARIMA and SARIMA indicate a steady increase in Vitamin D prescriptions over the next 5 years, highlighting the growing awareness and diagnosis of Vitamin D deficiency.

5.2 | Colecalciferol's Dominance in Prescriptions

The dashboard offers a comprehensive analysis of Vitamin D prescriptions over a period from April 2017 to December 2023, focusing on the use of three primary chemical substances: Colecalciferol, Alfacalcidol, and Ergocalciferol. Among these, Colecalciferol emerged as the most frequently prescribed, followed by Alfacalcidol, with Ergocalciferol being the least utilized, from Figure 9.

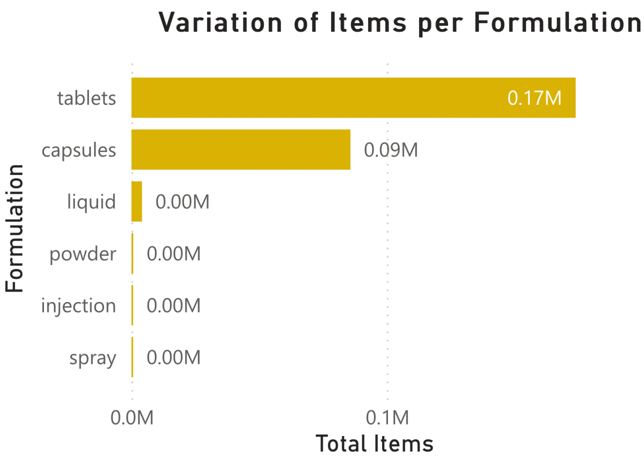

FIGURE 6 | Variation of items per formulation.

The analysis identified 60 different medications (Figure 10) and 74 distinct dosages (Figure 11) across six formulations, with Colecalciferol being available in the most diverse forms, particularly, in tablets, which were the most commonly prescribed. In contrast, Alfacalcidol was primarily available in capsules and liquid form, while Ergocalciferol, despite being available in multiple formulations, was prescribed less often. This preference for Colecalciferol is further reflected in the top medications, such as Accrete D3, where it was the main active ingredient.

5.3 | Dosage and Formulation Preferences

Additionally, we analyze dosage trends, revealing that unit-based dosages were the most common, particularly, in the 0–4999 range for both capsules and tablets. Calcium-based and milligram-based dosages were also frequently used, with specific ranges being more popular depending on the formulation in Figures 11 and 12. In Figure 13, the data show a consistent preference for tablets and capsules over other forms like liquids, powders, sprays, and injections across different surgeries. This trend underscores the dominant role of Colecalciferol-based formulations in Vitamin D prescriptions, which maintained high usage levels throughout the studied period. The detailed breakdown of dosages and formulations reflects the medical community's preference for specific Vitamin D forms and dosages to meet patient needs effectively.

5.4 | Technological Challenges

Implementing this dashboard involved several technical challenges:

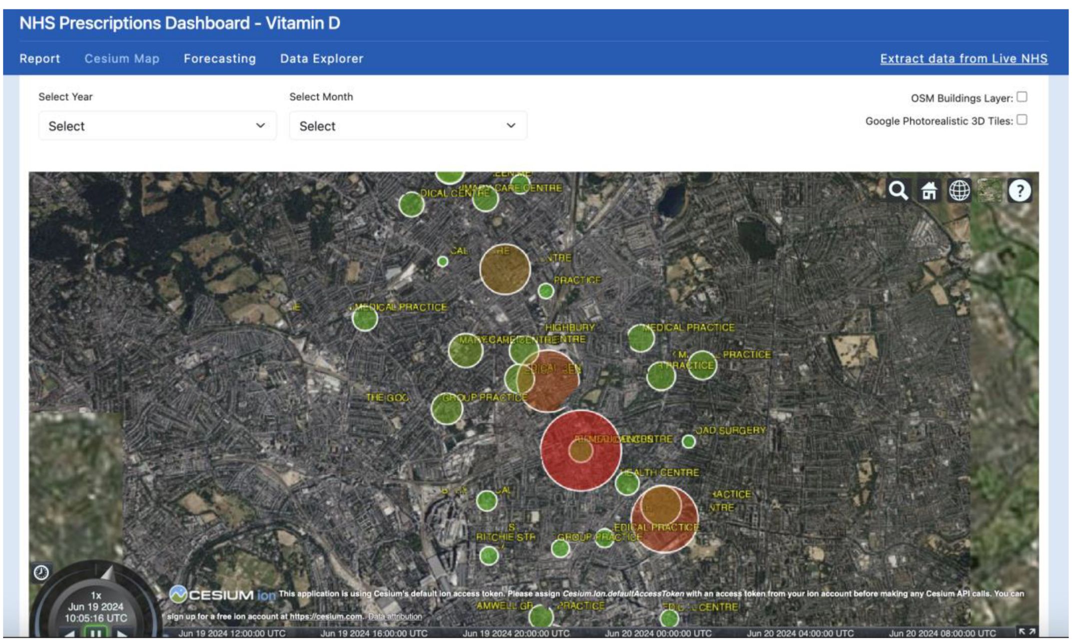

FIGURE 7 | NHS prescriptions dashboard—Vitamin D (Cesium map).

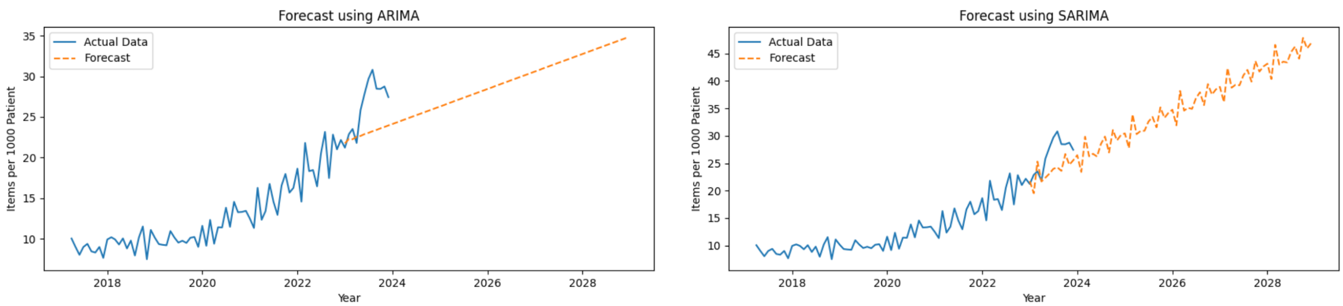

FIGURE 8 | ARIMA and SARIMA model forecasting.

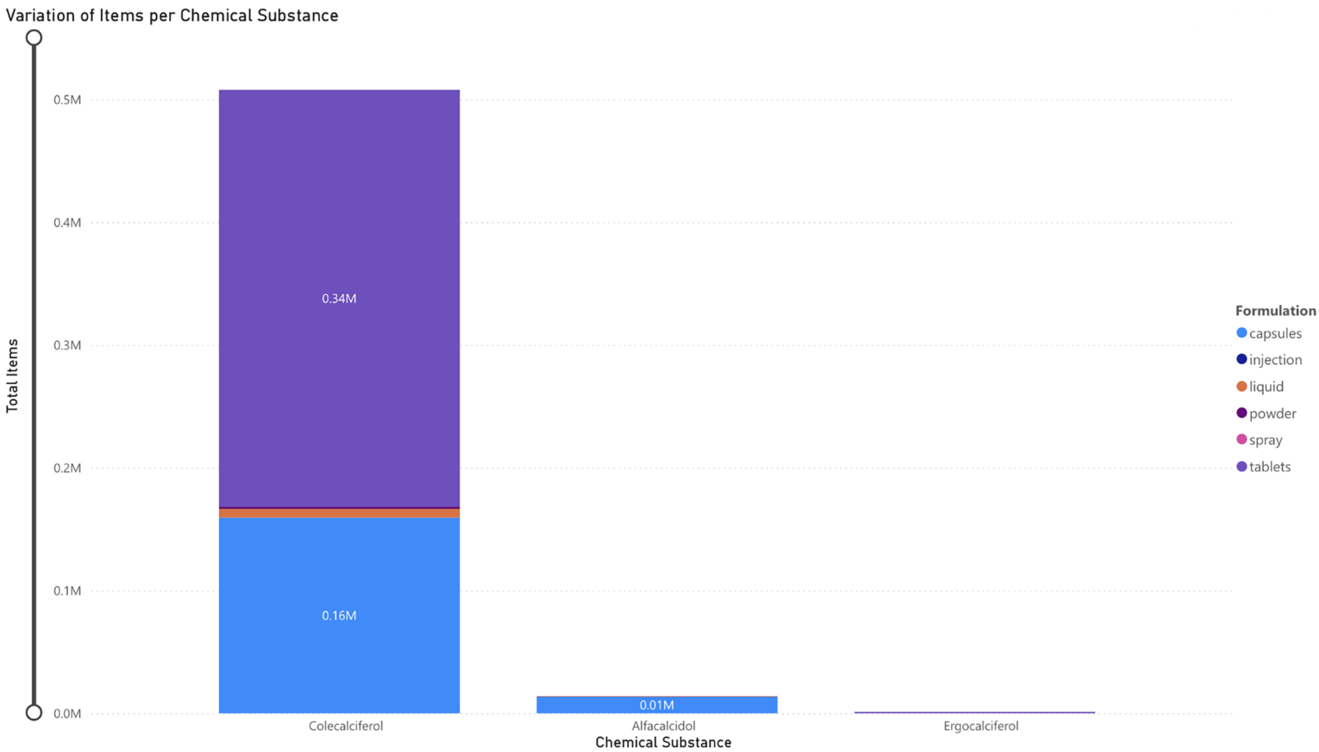

FIGURE 9 | Variation of items per chemical substances.

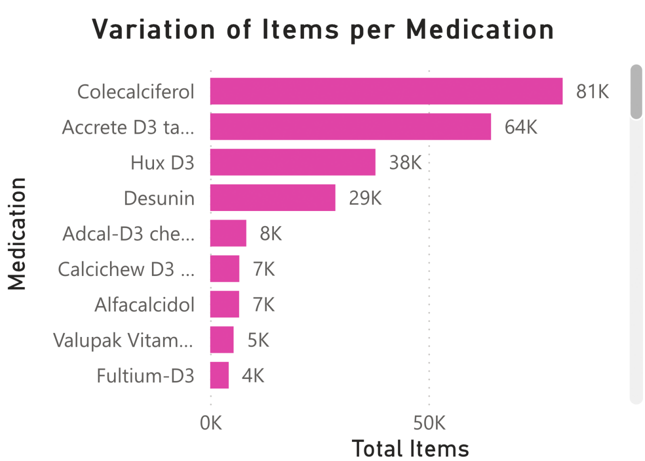

FIGURE 10 | Variation of items per formulation.

- Generalization of Patient Counts: Accurately generalizing patient counts using digital NHS website data was essential but challenging. Ensuring that patient data was correctly aggregated and anonymized while maintaining accuracy required sophisticated data-handling techniques.
- Compatibility Issues: Deploying PySpark in Azure Cloud presented difficulties due to the need for Java libraries, making pure Python solutions preferable for smaller datasets. While PySpark is highly effective for large-scale data processing, its deployment in Azure Cloud proved difficult and excessive for our needs.
- Data Manipulation: Ensuring accurate manipulation and categorization of the BNF description (medicine name, medication type, formulation, dosage type, and dosages) required meticulous data processing. This step was crucial for deriving meaningful insights from the data, but it involved complex text mining and data cleaning processes.

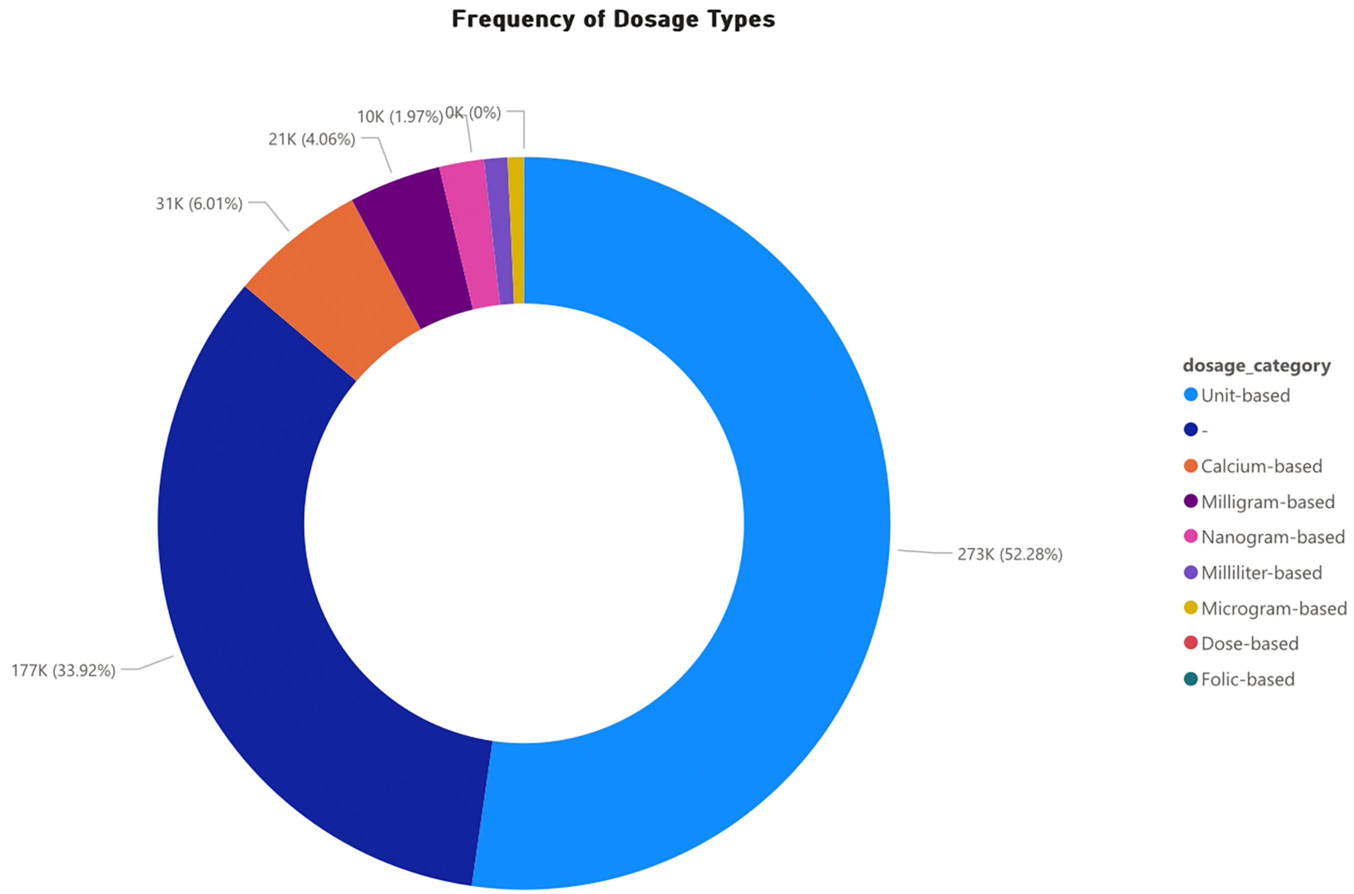

FIGURE 11 | Frequency of dosages.

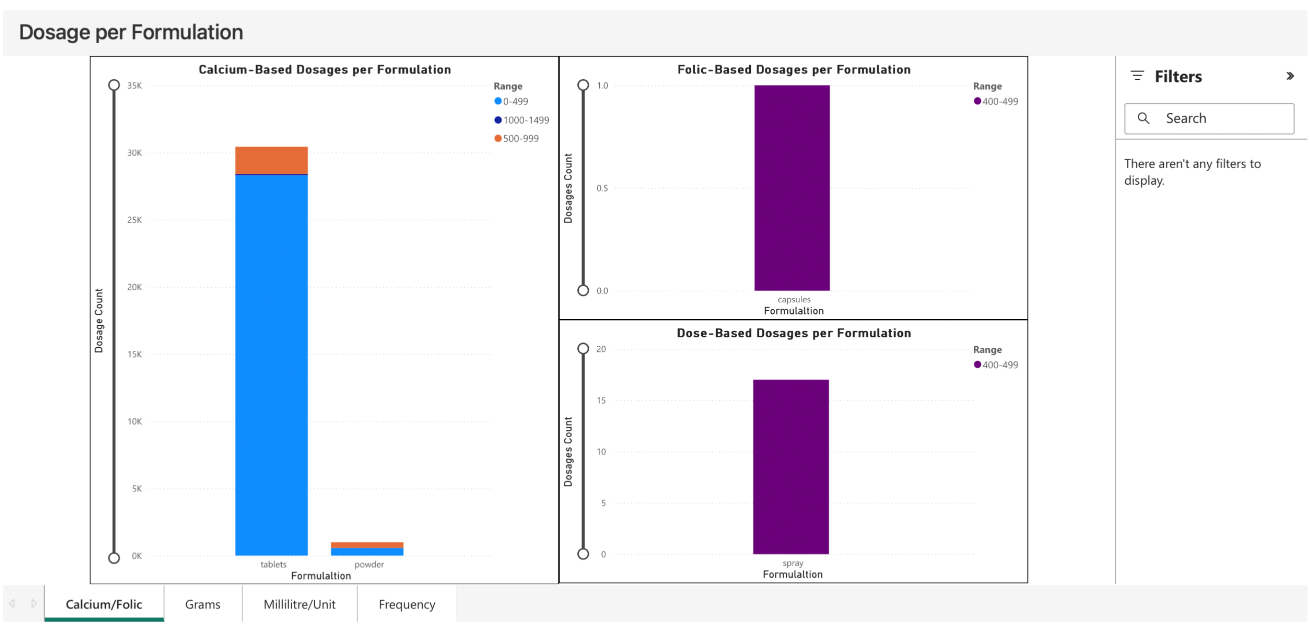

FIGURE 12 | Frequency of dosages.

- Integration with External Tools: Integrating PowerBI charts with the Flask application faced browser restrictions and login requirements, complicating external access. Ensuring seamless access and interaction for users outside the university domain required additional authentication and security measures.
- Scalability: While our current setup with Flask and Python libraries like Pandas and NumPy works well for the borough-specific dataset, scaling this system to handle national data would require transitioning to more robust data processing frameworks like Apache Spark. Spark's ability to efficiently process large datasets would be necessary to

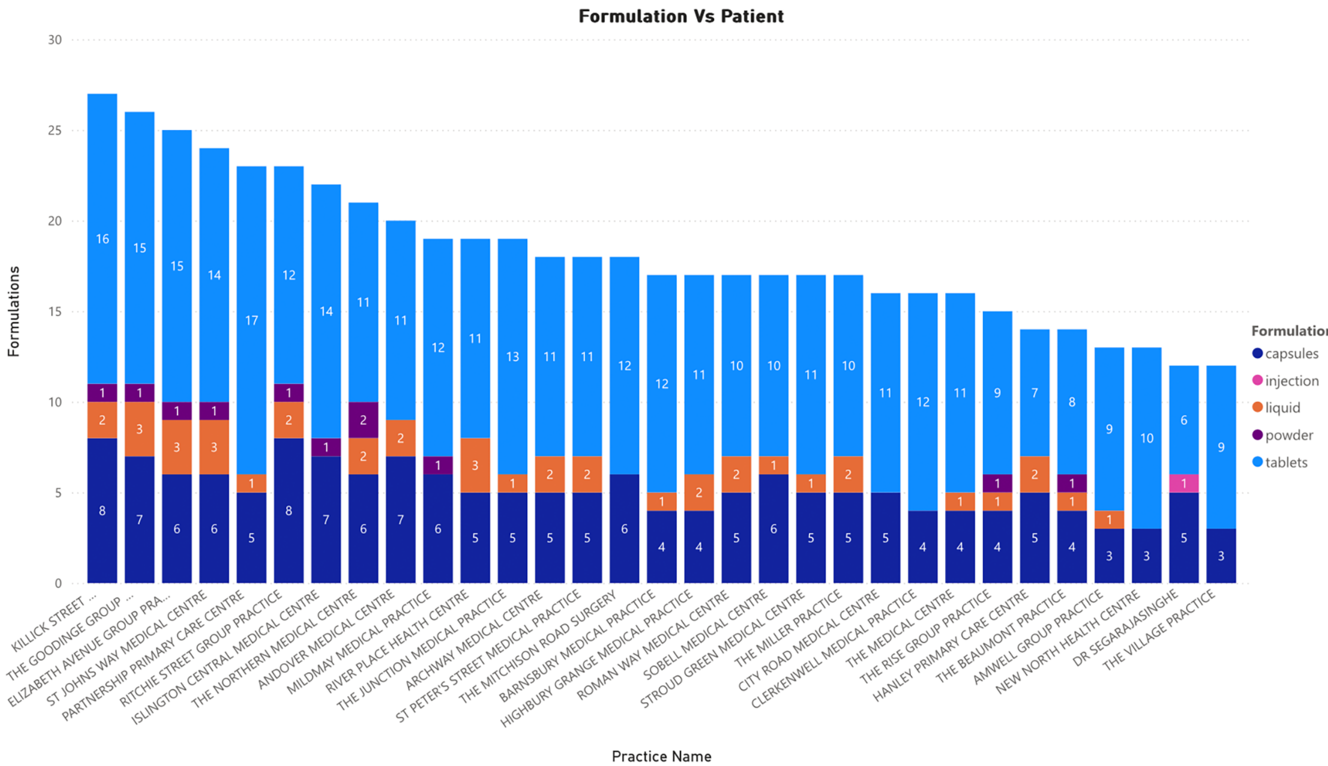

FIGURE 13 | Formulation per 1000 patients.

manage the increased volume and complexity of nationwide data.

Despite these challenges, the combination of Flask for the web application and Python libraries such as Pandas and NumPy for data processing has proven effective for our localized study with enhanced features that are not presented comprehensively in similar dashboards [4].

6 | Conclusion

We have developed a comprehensive dashboard to analyze Vitamin D prescription patterns of an inner London borough. Using Python libraries such as Pandas, NumPy, matplotlib, etc., we processed and visualized the data through an interactive Flask web application and deployed it in Azure Cloud.

The BNF_DESCRIPTION column was text-mined to extract medication, dosage, and formulation information. Formulations were categorized into liquid, powder, capsules, tablets, chewable tablets, injection, and spray. This categorization enabled the development of various charts and reports, including a "Quick Summary" section with multiple visualizations such as trends by year/month, variations by practice, and medication dosage, all at a glance. The data spans 7 years and allows for the addition of new, real-time data. Users can easily export data into formats like CSV or PDF directly from the system. Thirteen different raw data reports are available for viewing and downloading in CSV and PDF formats, along with eight different interactive reports with various filter options, and a quick summary designed in PowerBI. The Vitamin D dashboard provides an interactive 3D visualization (Cesium) of prescription data, enabling users to explore spatial trends and variations dynamically and intuitively. The Cesium offers a 3D visualization of all surgeries, with the size of the circle denoting patient count and the color indicating items prescribed per 1000 patients, ranging from green (low) to red (high). Additionally, OSM Buildings Layer and Google Photorealistic 3D Tiles options enhance the interactive user experience. Finally, a forecasting module predicts future prescription trends, supporting proactive decision-making using ARIMA and SARIMA models.

There are numerous benefits of our work. Primarily, the detailed visualization of prescription patterns and trends helps healthcare providers better understand the factors influencing Vitamin D deficiency and its treatment. This improved understanding allows for more informed, data-driven decision-making. Furthermore, the dashboard enables targeted interventions by identifying high-risk areas and populations, which can lead to more efficient resource allocation. Additionally, our predictive models offer foresight into future prescription needs, aiding strategic planning and policy formulation. Finally, the web-based nature of the dashboard enhances accessibility for stakeholders, fostering transparency, and collaboration within the healthcare community.

The results from our analysis can be used in several impactful ways. For instance, they can inform healthcare policies adapted to address Vitamin D deficiency more effectively. By targeting interventions based on identified high-risk areas and populations, healthcare providers can conduct educational campaigns and health initiatives where they are most needed. Furthermore, the ability to forecast prescription trends allows for optimal resource allocation, ensuring that supplies meet future demands. Ultimately, these insights contribute to enhancing patient care by addressing the underlying causes of Vitamin D deficiency more comprehensively.

Despite the significant advancements achieved, there are several areas for additional enhancement. One crucial next step is to expand the scope of the analysis to include data from across the entire United Kingdom. This broader dataset will provide a more comprehensive understanding of national trends in Vitamin D prescriptions. Another important step is to integrate additional datasets, such as patient demographics and health outcomes, which will enrich the analysis and provide a more holistic view of the factors affecting Vitamin D deficiency. Additionally, refining the predictive models to increase their accuracy and reliability is vital. Finally, providing training for healthcare providers and stakeholders on effectively using the dashboard will maximize its utility and impact.

To accomplish these goals, we propose several actions. Transitioning to Apache Spark for data processing will be fundamental for handling the larger datasets involved in a nationwide analysis. Collaborating with data providers to access and integrate additional relevant datasets will enhance the depth and breadth of our analysis. Continuously updating and validating predictive models using the latest data and advanced machine-learning techniques will ensure the models remain accurate and reliable. Engaging stakeholders through workshops and training sessions will facilitate effective use of the dashboard and gather valuable feedback for further improvements.

By following these steps, we can build on the foundation of our current work to provide even more valuable insights and support for healthcare providers in managing Vitamin D deficiency. This expanded and enhanced approach will help to address Vitamin D deficiency more effectively at both local and national levels, ultimately leading to better health outcomes for the population.

Data Availability Statement

The data that support the findings of this study are available in NHS prescription data at https://www.nhsbsa.nhs.uk/prescription-data/prescribing-data/english-prescribing-data-epd. These data were derived from the following resources available in the public domain: prescription-data, https://www.nhsbsa.nhs.uk/prescription-data/prescribing-data/english-prescribing-data-epd.

References

1. M. F. Holick, "Vitamin D Deficiency," *New England Journal of Medicine* 357 (2007): 266–281.

2. U. Fairbrother, "Telomere Oxidation Status (TOS) Is Correlated With Relative Telomere Length (RTL) Across Different Mouse Tissues but Not With Nicotinamide Nucleotide Transhydrogenase (Nnt) Status," 2023, accessed June 13, 2024, repository.londonmet.ac.uk.

3. Nordic IT Project, "Igloo Reader—Point of Care Analyzer," accessed January 4, 2025, https://nordicitproject.com/igloo-reader.

4. OpenPrescribing.net, "Bennett Institute for Applied Data Science, University of Oxford," 2024, accessed August 13, 2024, https://openprescribing.net/.

5. NHS Business Services Authority, "ePACT2 Overview" 2024, accessed August 22, 2024, https://www.nhsbsa.nhs.uk/epact2.

6. OpenSAFELY, "Research Platform Overview," 2024, accessed August 22, 2024, https://www.opensafely.org/.

7. PrescQIPP, "About Us," 2024, accessed August 22, 2024, https://www.prescqipp.info/.

8. University of Oxford, "OpenPrescribing: Putting Data and Statistics Into Action to Save the NHS Money" 2024, accessed August 13, 2024, https://www.ox.ac.uk/news/2015-12-07-nhs-gp-prescribing-data-open-all.

9. HealthData.gov, *U.S. Department of Health & Human Services* (HealthData.gov, 2024).

10. Institute for Health Metrics and Evaluation (IHME), "Interactive Data Visuals," accessed August 13, 2024, https://www.healthdata.org/data-tools-practices/interactive-data-visuals.

11. Flare Compare, "Plotly vs. D3.js," 2021, accessed August 13, 2024, https://flarecompare.com/Data%20Visualization/Plotly%20vs.%20D3.js.

12. B. W. Hollis, D. Johnson, T. C. Hulsey, M. Ebeling, and C. L. Wagner, "Vitamin D Supplementation During Pregnancy: Double-Blind, Randomized Clinical Trial of Safety and Effectiveness," *Journal of Bone and Mineral Research* 26, no. 10 (2011): 2341–2357.

13. A. Martinea, C. Camargo, K. S. Khan, and R. Hooper, "Vitamin D Supplementation to Prevent Acute Respiratory Infections: Systematic Review and Meta-Analysis of Individual Participant Data," *BMJ* 356 (2019): i6583, https://fundingawards.nihr.ac.uk/award/13/03/25.

14. J. E. Manson, N. R. Cook, I. M. Lee, et al., "Vitamin D Supplements and Prevention of Cancer and Cardiovascular Disease," *New England Journal of Medicine* 380, no. 1 (2019): 33–44.

15. H. A. Bischoff-Ferrari, B. Dawson-Hughes, H. B. Staehelin, et al., "Fall Prevention With Supplemental and Active Forms of Vitamin D: A Meta-Analysis of Randomised Controlled Trials," *BMJ* 339 (2009): b3692.

16. S. Kumar, A. Haleem, and R. P. Singh, "Health Data Portals and Their Impact on Public Health," *Journal of Health Informatics* 14, no. 3 (2021): 120–135.

17. Dask, "Dask: Parallel Computing With Task Scheduling," 2024 Dask.org, accessed August 22, 2024, https://dask.org.