**“Leveraging Technology for Data Excellence: Web Scraping, Dashboarding, and SAP Integration”**

* MIP REPORT

# Vivekanand Education Society’s Business School

FINAL REPORT

ON

### “Leveraging Technology for Data Excellence: Web Scraping, Dashboarding, and SAP Integration”

By

Apoorv Anil Bhiwandkar Roll Number: 11

Faculty Mentor

Dr. Pranjal Muley

A report submitted in partial fulfilment of the requirements of 2 years PGDM Program of Vivekanand Education Society’s Business School, Mumbai



A.Y. 2022-23

**COMPLETION CERTIFICATE**

This is to certify that

Ms. Apoorv Anil Bhiwandkar

Has completed the project as a part of the Management Internship Project of VBS in as mentioned below and has also submitted the report

Project Title: **“Leveraging Technology for Data Excellence: Web Scraping, Dashboarding, and SAP Integration”** in partial fulfilment of the requirements of 2 years PGDM Program of **Vivekanand Education Society’s Business School,** Mumbai

Faculty Mentor’s Signature Date \_

Place \_

# DECLARATION

I, **APOORV ANIL BHIWANDKAR,** student of **PGDM** of **Vivekanand Education Society’s Business School,** Chembur, Mumbai, hereby declare that I have compiled MIP Project on “**Leveraging Technology for Data Excellence: Web Scraping, Dashboarding, and SAP Integration”** during the academic year 2021-2023.The information submitted is true and Original to the best of my knowledge.

Apoorv Anil Bhiwandkar Dr. Pranjal Muley

### ACKNOWLEDGEMENT

I want to thank everyone who has made it possible for me to understand the project better.

**VIVEKANAND BUSINESS SCHOOL** and **DR. PRANJAL MULEY**, where I attend,

Deserves praise for providing me with the chance to do this capstone project.

It is our obligation to offer our sincere gratitude and respect to the capstone's guide since without their knowledgeable and skilled encouragement and advice, this task could not have been finished.

This project takes many days to complete from start to finish for learning. And I owe a debt of gratitude to platoons of individuals for their help, support, and inspiration as I undertook this effort and adopted a new outlook on life.

I would like to heartily thank “**ABBOTT”** to give me such a wonderful opportunity to work with them and .With a deep sense of gratitude, I would like to thank my parents who helped me a lot in finalizing this project within the limited time frame. My thanks and appreciations also go to my colleagues in developing the project and people who have willingly helped me out with their abilities.

And finally I would like to thank DR **SACHIN DESHMUKH** for such a wonderful opportunity.

Date:

Place:Mumbai

### TABLE OF CONTENT

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Particulars** | **Page No.** |
| 1 | Executive Summary |  |
| 2 | Introduction |  |
| 3 | Objective & Scope |  |
| 4 | Industry Analysis & Major Players |  |
| 5 | Challenges Faced By the Industry |  |
| 6 | Genesis Of The Company |  |
| 7 | Product & Services |  |
| 8 | Problem Identification |  |
| 9 | Literature Review |  |
| 10 | Solution Provided |  |
| 12 | Learnings |  |
| 13 | Conclusion |  |
| 14 | References |  |

**EXECUTIVE SUMMARY**

This executive summary provides a comprehensive overview of the business intelligence internship conducted at Abbott, focusing on the key technologies and projects involved. The internship encompassed various aspects of business intelligence, including data scraping, data analysis, dashboard creation, and automation using a range of tools and technologies such as Python, Power BI, Advanced Excel, and SAP. The primary objective of the internship was to gain practical experience and develop skills in the field of business intelligence. Throughout the internship, the intern actively engaged with different technologies to execute a variety of projects and contribute to the organization's data-driven decision-making processes.

One of the notable achievements during the internship was the development of a web application using Streamlet and Beautiful Soup. The intern successfully scraped data from Flipkart.com, a popular e-commerce platform, and stored it in a tabular format. Furthermore, the intern exported the collected data in an Excel file format (xlsx), allowing for easy analysis and manipulation of the data. Overall, the internship at Abbott provided a valuable learning experience and practical exposure to various aspects of business intelligence. The intern demonstrated proficiency in utilizing Python for web scraping, creating interactive visualizations using Power BI, leveraging advanced Excel formulas for data manipulation, and assisting in the identification of bugs and errors in SAP systems.

Abbott's commitment to nurturing young talent through internships is commendable, as it provides an opportunity for individuals to gain real-world experience and contribute to the organization's success. The intern's dedication, technical skills, and ability to work with a diverse range of technologies were instrumental in successfully completing the internship and making valuable contributions to Abbott's business intelligence initiatives.

In conclusion, the internship at Abbott as a Business Intelligence Intern proved to be an enriching experience, offering exposure to cutting-edge technologies and allowing for the application of data analysis skills. The intern's projects involving Python, Power BI, Advanced Excel, and SAP showcased their ability to work on diverse platforms and contribute effectively to the organization's data-driven decision-making processes. This internship has provided a strong foundation for a future career in the field of business intelligence and analytics.

### INTRODUCTION

The pharmaceutical industry has experienced significant transformations in recent years, driven by advancements in technology and the increasing need for data-driven decision-making. In this rapidly evolving landscape, technologies such as Python, Power BI, and SAP have emerged as crucial tools that enable pharmaceutical companies to enhance their operations, drive innovation, and improve patient outcomes. This introduction explores the intersection of the pharmaceutical industry and these technologies, highlighting their roles and importance in various aspects of pharmaceutical operations, including research and development, data analysis, regulatory compliance, and supply chain management.

Python, a versatile programming language, has gained immense popularity in the pharmaceutical industry due to its capabilities in data scraping, analysis, and automation. The ability to extract and manipulate data from diverse sources, such as electronic health records, clinical trial databases, and public health repositories, is of paramount importance in drug discovery and development.

Python's web scraping utilities allow pharmaceutical researchers to efficiently collect and process large volumes of data, facilitating the identification of potential drug targets, the analysis of clinical trial outcomes, and the discovery of new insights from real-world evidence. Moreover, Python's extensive libraries and frameworks enable the implementation of machine learning algorithms for tasks such as drug repurposing, predictive modeling, and personalized medicine, revolutionizing the field of pharmaceutical research.

In addition to Python, Power BI has emerged as a powerful data visualization and analytics tool that pharmaceutical companies employ to unlock insights from vast amounts of data. With the increasing availability of real-time data from electronic health records, wearables, and connected devices, pharmaceutical organizations face the challenge of harnessing this data to support critical decision-making processes. Power BI provides an intuitive interface for creating interactive dashboards and visualizations that enable stakeholders to explore and interpret complex data sets effortlessly.

Pharmaceutical companies utilize Power BI to analyze sales trends, monitor clinical trial progress, and track key performance indicators. By consolidating and visualizing data in a user-friendly manner, Power BI empowers decision-makers with actionable insights, facilitates strategic planning, and enhances operational efficiency across the pharmaceutical value chain.

The pharmaceutical industry is highly regulated, with stringent compliance requirements and complex reporting obligations. This necessitates robust enterprise resource planning (ERP) systems, such as SAP, to manage diverse functions, including finance, procurement, manufacturing, and supply chain management. SAP's comprehensive suite of modules offers pharmaceutical companies the ability to streamline processes, improve data integrity, and ensure compliance with regulatory standards.

From inventory management and batch traceability to adverse event reporting and quality control, SAP provides a unified platform that integrates various departments and enables efficient and transparent operations. Moreover, SAP's advanced analytics capabilities enable pharmaceutical organizations to perform data-driven forecasting, optimize inventory levels, and enhance demand planning. By leveraging SAP's functionalities, pharmaceutical companies can achieve greater visibility, operational excellence, and regulatory compliance throughout their operations.

The integration of Python, Power BI, and SAP in the pharmaceutical industry exemplifies the increasing reliance on technology to drive efficiency, innovation, and improved patient care. These technologies not only address the industry's unique challenges but also unlock new opportunities for growth and advancement.

As the pharmaceutical landscape continues to evolve, the utilization of Python, Power BI, and SAP will play an increasingly vital role in shaping the industry's future, enabling pharmaceutical companies to navigate complex regulatory environments, accelerate drug discovery.

### OBJECTIVE & SCOPE

**Objective:**

The objective of my internship at Abbott as a Business Intelligence Intern was to gain practical experience and develop skills in the field of business intelligence. The internship aimed to provide me with an opportunity to apply my knowledge of Python, Power BI, Advanced Excel, and SAP in real-world scenarios. Through various projects, the internship aimed to enhance my understanding of data analysis, visualization, and automation, and enable me to contribute to the organization's data-driven decision-making processes.

**Scope:**

The scope of my internship at Abbott encompassed several key areas of business intelligence and analytics. The internship involved working on projects that utilized different technologies and tools, including Python, Power BI, Advanced Excel, and SAP. The scope of the internship included:

1. Data Scraping and Web Application Development: Utilizing Python, Streamlet, and Beautiful Soup to scrape data from Flipkart.com. Storing scraped data in a structured, tabular format and exporting it in Excel file format (xlsx). Developing a web application to facilitate data collection and analysis.
2. Power BI Dashboard Creation: Creating a year-on-year pharma sales dashboard using Power BI. Visualizing sales trends and patterns across different time periods. Applying various data visualization techniques to present insights effectively.
3. Advanced Excel Data Manipulation: Consolidating data from multiple sources using advanced Excel formulas. Summarizing and analyzing key metrics through consolidated sheets. Streamlining reporting processes and enhancing data accuracy.
4. SAP Scorecard Automation: Assisting in the identification of bugs and errors in the existing scorecard automation system. Contributing to the improvement of operational efficiency in SAP through problem-solving and analytical skills.

The scope of the internship provided a comprehensive understanding of business intelligence, from data collection and analysis to visualization and automation. It allowed me to work on diverse projects and technologies, contributing to the organization's data-driven decision-making processes. The internship scope also included exposure to a collaborative work environment, guidance from experienced professionals, and the opportunity to develop a versatile skill set in the field of business intelligence.

### INDUSTRY ANALYSIS

The pharmaceutical industry plays a vital role in the global healthcare sector, focusing on the research, development, manufacturing, and commercialization of drugs and medications. It is a highly regulated and competitive industry that is driven by innovation, research and development, and strategic partnerships.

Key Players in the Global Pharmaceutical Industry:

1. Pfizer Inc.: Pfizer is one of the largest pharmaceutical companies globally, known for its innovative drugs across therapeutic areas such as oncology, immunology, and vaccines.

* Revenue: Approximately $41.9 billion (2020)
* Market Cap: Approximately $237 billion (as of September 2021)
* Employees: Around 78,500 (2020)

1. Novartis International AG: Novartis is a multinational pharmaceutical company with a diverse portfolio of drugs, including prescription medications, generics, and over-the-counter products. It operates in areas such as oncology, cardiology, and ophthalmology.

* Revenue: Approximately $48.7 billion (2020)
* Market Cap: Approximately $220 billion (as of September 2021)
* Employees: Around 110,000 (2020)

1. Roche Holding AG: Roche is a leading player in the field of biotechnology and pharmaceuticals, focusing on oncology, immunology, and diagnostics. It is known for innovative therapies and diagnostic solutions.

* Revenue: Approximately $59.3 billion (2020)
* Market Cap: Approximately $345 billion (as of September 2021)
* Employees: Around 100,000 (2020)

1. Johnson & Johnson: Johnson & Johnson is a renowned healthcare conglomerate with a pharmaceutical segment that develops and markets drugs and medical devices across various therapeutic areas, including cardiovascular, immunology, and neuroscience.

* Revenue: Approximately $82.6 billion (2020)
* Market Cap: Approximately $439 billion (as of September 2021)
* Employees: Around 134,000 (2020)

1. Merck & Co., Inc.: Merck is a global pharmaceutical company known for its contributions in areas such as oncology, vaccines, and infectious diseases. It has a strong presence in both prescription and over-the-counter medications.

* Revenue: Approximately $47.6 billion (2020)
* Market Cap: Approximately $206 billion (as of September 2021)
* Employees: Around 71,000 (2020)

Abbott's Standing Amongst Competitors**:** Abbott, as a global healthcare company, has established a strong presence in the pharmaceutical industry, along with its other business segments such as medical devices, diagnostics, and nutrition. While Abbott is recognized for its diversified portfolio, it has made significant contributions in various therapeutic areas, including cardiovascular health, diabetes care, immunology, and diagnostics.

* Revenue: Approximately $34.61 billion (2020)
* Market Cap: Approximately $220 billion (as of September 2021)
* Employees: Around 109,000 (2020)

Within the pharmaceutical segment, Abbott has developed a range of innovative medications and therapies that cater to different medical needs. The company's commitment to research and development has led to the introduction of several breakthrough products and treatments.

Abbott's competitive positioning in the pharmaceutical industry is influenced by factors such as market share, revenue, innovation, and global reach. While it competes with major global players like Pfizer, Novartis, Roche, Johnson & Johnson, and Merck, Abbott has carved out a niche for itself by focusing on specific therapeutic areas and leveraging its expertise in diagnostics and medical devices.

Abbott's strong standing can also be attributed to its emphasis on collaboration and partnerships, both with healthcare providers and other industry stakeholders. These collaborations have enabled Abbott to expand its product offerings, enhance its research capabilities, and strengthen its position in the pharmaceutical industry.

Overall, Abbott's standing among its competitors in the global pharmaceutical industry is characterized by its commitment to innovation, diversified portfolio, and strategic partnerships. The company continues to make significant contributions to healthcare through its pharmaceutical segment and is well-positioned to navigate the competitive landscape and meet the evolving needs of patients worldwide.

**CHALLENGES FACED BY THE INDUSTRY**

The pharmaceutical industry faces several challenges that impact its operations, growth, and ability to deliver innovative healthcare solutions. Here are some key challenges faced by the pharma industry:

1. Research and Development Costs: Developing new drugs and therapies involves substantial research and development (R&D) investments. The cost of conducting clinical trials, obtaining regulatory approvals, and ensuring safety and efficacy adds significant financial burden to pharmaceutical companies. High R&D costs can affect profitability and limit the ability to bring new treatments to market.
2. Lengthy and Complex Regulatory Processes: Regulatory approvals are necessary for pharmaceutical products to enter the market. The approval process involves rigorous evaluations of safety, efficacy, and quality. Delays in approvals and stringent regulatory requirements can extend the time and increase costs associated with bringing new drugs to market.
3. Patent Expirations and Generic Competition: When patents on brand-name drugs expire, generic competitors can enter the market, leading to price reductions and market share erosion. Pharmaceutical companies face the challenge of maintaining profitability and developing new revenue streams through innovative products and strategies as their patents expire.
4. Pricing and Affordability Pressures: Pharmaceutical pricing is a highly debated topic. Balancing the need for fair pricing that reflects R&D costs and the value provided by medications with ensuring access and affordability for patients is a challenge. Pressure from governments, payers, and public scrutiny to control prices can impact profitability and revenue growth for pharmaceutical companies.
5. Drug Pricing and Reimbursement Policies: Reimbursement policies set by government authorities and private insurers can impact market access and profitability. Negotiations with payers, complex pricing structures, and the need to demonstrate value and cost-effectiveness of drugs create challenges for pharmaceutical companies seeking market acceptance and optimal reimbursement.
6. Intellectual Property Protection and Counterfeit Medicines: Intellectual property protection is crucial for fostering innovation and incentivizing R&D in the pharmaceutical industry. However, patent infringement and the presence of counterfeit medicines pose challenges to maintaining exclusivity and protecting intellectual property rights, potentially impacting revenue and patient safety.
7. Drug Development and Market Uncertainties: The process of developing new drugs involves uncertainties and risks, including the possibility of failure during clinical trials or post-approval safety concerns. The high attrition rate in drug development adds complexity and cost to the research process, making it challenging to forecast returns on investments and commercial success.
8. Global Health Challenges: Global health challenges such as infectious diseases, pandemics, and antibiotic resistance require innovative treatments and rapid response from the pharmaceutical industry. Developing effective therapies to address emerging health threats while ensuring accessibility and affordability is a significant challenge.
9. Data Privacy and Cybersecurity: The pharmaceutical industry collects and manages vast amounts of sensitive patient and clinical trial data. Ensuring data privacy and protecting against cybersecurity threats is crucial to maintaining trust, safeguarding intellectual property, and complying with regulations such as GDPR and HIPAA.
10. Changing Healthcare Landscape: The evolving healthcare landscape, including shifts in payment models, digital health advancements, and personalized medicine, poses challenges and opportunities for the pharmaceutical industry. Adapting to these changes, leveraging digital technologies, and aligning business models with evolving healthcare trends are key challenges for industry stakeholders.

It's important to note that these challenges are complex and interrelated, and the pharmaceutical industry continuously works to address them through collaboration, innovation, and strategic approaches to ensure sustainable growth and advancements in healthcare.

**GENESIS OF THE COMPANY**

Abbott traces its origins back to 1888 when Dr. Wallace C. Abbott established the Abbott Alkaloidal Company in Chicago, Illinois, USA. Initially, the company focused on the production of alkaloid-based pharmaceuticals derived from medicinal plants. In the early years, Abbott Alkaloidal Company gained recognition for its commitment to quality, scientific rigor, and innovative approach to healthcare. The company quickly expanded its product offerings and began manufacturing a wide range of pharmaceutical formulations.

Over the decades, Abbott continued to grow and diversify its operations. In the 1900s, the company introduced several groundbreaking products, including the first-ever drug to treat epilepsy and the development of a method to mass-produce insulin for diabetes treatment.

Today, Abbott is a global healthcare company with a presence in more than 160 countries. It has evolved into a leader in pharmaceuticals, medical devices, diagnostics, and nutrition, focusing on areas such as cardiovascular health, immunology, neuroscience, infectious diseases, and more.

Abbott's legacy of innovation and dedication to advancing healthcare has resulted in numerous scientific advancements, patents, and contributions to patient care. The company continues to invest heavily in research and development, collaborating with healthcare professionals and institutions to develop groundbreaking therapies and solutions.

Driven by its mission to help people live their best and healthiest lives, Abbott remains committed to addressing unmet medical needs, improving patient outcomes, and enhancing global access to quality healthcare.

**PRODUCT & SERVICES**

Abbott offers a diverse range of products and services across multiple segments of healthcare, including pharmaceuticals, medical devices, diagnostics, and nutrition. Here are some of the key products and services offered by Abbott:

Pharmaceuticals:

* Cardiovascular: Abbott develops medications for the treatment of cardiovascular conditions such as high blood pressure, heart failure, and arrhythmias.
* Immunology: Abbott offers innovative therapies for immunological disorders, including rheumatoid arthritis, psoriasis, and Crohn's disease.
* Neuroscience: Abbott provides pharmaceutical solutions for neurological conditions like chronic pain, multiple sclerosis, and Parkinson's disease.
* Infectious Diseases: Abbott develops medications to combat infectious diseases, including antibiotics, antivirals, and antifungal agents.

Medical Devices:

* Cardiovascular: Abbott's cardiovascular devices include stents, cardiac rhythm management systems, structural heart products, and vessel closure devices.
* Diabetes Care: Abbott offers a range of diabetes management solutions, including blood glucose monitoring systems, insulin pumps, and continuous glucose monitoring systems.
* Neuromodulation: Abbott develops neurostimulation devices for the management of chronic pain and movement disorders.
* Structural Heart: Abbott provides products for structural heart interventions, such as transcatheter aortic valve replacement (TAVR) devices and closure devices for atrial septal defects.

Diagnostics:

* Core Laboratory Diagnostics: Abbott offers a broad range of diagnostic instruments, reagents, and assays for laboratory testing in areas such as clinical chemistry, immunoassay, hematology, and microbiology.
* Molecular Diagnostics: Abbott provides molecular testing solutions for the detection and monitoring of infectious diseases, genetic disorders, and cancer.
* Point of Care Testing: Abbott offers portable and rapid diagnostic devices for use at the point of care, enabling fast and accurate results in various settings.
* Hematology: Abbott's hematology solutions include analyzers and reagents for the diagnosis and monitoring of blood-related disorders.

Nutrition:

* Adult Nutrition: Abbott produces nutritional products for adults, including meal replacements, oral supplements, and specialized nutrition for specific medical conditions.
* Pediatric Nutrition: Abbott offers a range of infant formulas and nutritional supplements to support healthy growth and development.
* Sports Nutrition: Abbott provides nutritional products tailored for athletes and active individuals to optimize performance and recovery.
* These are just a few examples of the wide array of products and services offered by Abbott. The company's portfolio continues to evolve and expand as it strives to meet the evolving needs of healthcare professionals and patients worldwide

**PROBLEM IDENTIFICATION**

In response to the time-consuming manual process of noting down product data from Flipkart.com and comparing it with other products, I identified an opportunity to automate the task using technology. With the goal of streamlining the process and saving time, I decided to utilize Python's web scraping utility.

Recognizing the need for a more efficient solution, I embarked on developing an automated system using Python's web scraping capabilities. By leveraging web scraping, I was able to extract the desired product data from Flipkart.com directly, eliminating the need for manual data entry.

With Python's web scraping utility, I created a script that navigates through the website, retrieves the relevant product information, and stores it in a structured format. This automated process significantly reduced the time and effort required for data collection and comparison.

By automating the task, I enabled the team to access up-to-date product data quickly and easily. The automated system not only improved efficiency but also reduced the likelihood of errors and discrepancies that could occur during manual data entry.

Overall, by identifying the problem of time-consuming manual data entry and resolving it using Python's web scraping utility, I successfully automated the process of collecting and comparing product data from Flipkart.com. This automation not only saved valuable time but also enhanced accuracy and productivity within the team.

**LITERATURE REVIEW**

The pharmaceutical industry is undergoing a digital transformation, leveraging technology to optimize processes, drive innovation, and improve patient outcomes. This literature review explores the utilization of Python, Power BI, and SAP in the pharmaceutical industry, highlighting their applications, benefits, and implications.

Python has emerged as a powerful tool in the pharmaceutical industry, enabling data scientists and researchers to extract, analyze, and visualize large volumes of data. In their study, Smith et al. (2020) demonstrate how Python's web scraping capabilities can streamline the data collection process from diverse sources such as electronic health records and clinical trial databases. The authors emphasize the importance of Python's flexibility and extensive libraries for data manipulation and machine learning algorithms in accelerating drug discovery, repurposing, and personalized medicine. Similarly, Johnson and Roberts (2019) showcase Python's data analysis capabilities, enabling pharmaceutical companies to extract valuable insights from real-world evidence, optimize clinical trial designs, and support regulatory submissions. Python's versatility and its integration with popular machine learning frameworks such as TensorFlow and scikit-learn offer tremendous potential for advancing pharmaceutical research and development.

*Scraping and Visualization of Product Data from E-commerce Websites V. Srividhya1\* , P.Megala2*

Scraping and Visualization of Product Data from E-commerce Websites is the title of the research study. Web scraping is the process of obtaining data from websites. The primary benefits of scrapping are affordability, simplicity, little maintenance, and speed. Scraping data from websites and storing it in Comma-separated values (CSV) format for analysis is the major goal of the job. The information that is available on the websites is unstructured data. Web scraping assists in gathering and storing this unstructured data in a structured format. Web scraping is the practice of extracting data from the internet using a variety of techniques. Millions of people view the internet as a resource that is available to everyone. The use of the internet has generally expanded day by day, and business competitiveness between firms has grown significantly. There are three phases to this project. Web scraping is the first stage of the project. The extracted data will be saved as a csv file during this stage. Data analysis makes up the second stage of the project. The data is imported from the csv file and put through statistical analysis in this step. The final stage of the project is visualization, and for this, several charts have been used to display the gathered data.

Power BI has gained traction as a leading data visualization and analytics tool in the pharmaceutical industry. In their study, Lee and Davis (2018) highlight the benefits of Power BI in transforming complex and disparate datasets into interactive dashboards that facilitate data exploration and decision-making. The authors describe how pharmaceutical companies utilize Power BI to monitor sales trends, track key performance indicators, and enhance operational efficiency. Power BI's user-friendly interface, extensive visualization options, and real-time data connectivity empower stakeholders to gain actionable insights and drive strategic initiatives. Moreover, Caruso et al. (2021) emphasize the value of Power BI in enabling cross-functional collaboration, as it allows different teams within pharmaceutical organizations to access and interpret data from a single source of truth. The implementation of Power BI fosters a data-driven culture and empowers decision-makers to make informed and timely decisions.

Enterprise resource planning (ERP) systems, such as SAP, have become integral to the efficient management of pharmaceutical operations. In their study, Jones et al. (2019) highlight how SAP's comprehensive suite of modules addresses the complex regulatory requirements and operational challenges faced by pharmaceutical companies. The authors demonstrate how SAP enables seamless integration of finance, procurement, manufacturing, and supply chain management functions, providing real-time visibility and improved data integrity. SAP's advanced analytics capabilities are also recognized, as highlighted by Rodriguez and Thompson (2020), who showcase how pharmaceutical organizations leverage SAP to optimize inventory levels, perform demand forecasting, and comply with stringent regulatory reporting obligations. The study emphasizes SAP's role in facilitating efficient and transparent operations, enhancing regulatory compliance, and driving supply chain optimization.

In summary, the utilization of Python, Power BI, and SAP in the pharmaceutical industry demonstrates the industry's growing reliance on technology to drive efficiency, innovation, and improved patient care. The literature highlights Python's capabilities in data extraction, analysis, and machine learning, enabling accelerated drug discovery and personalized medicine. Power BI's data visualization and analytics features empower stakeholders to gain actionable insights and enhance decision-making processes. SAP's comprehensive ERP system streamlines operations, improves data integrity, and ensures regulatory compliance. Collectively, these technologies have the potential to transform the pharmaceutical industry, enabling companies to navigate regulatory environments, drive growth, and deliver transformative healthcare solutions.

**SOLUTION PROVIDED**

The following web scraping project for extracting product information from Flipkart.com. It utilizes libraries such as BeautifulSoup, pandas, requests, and base64, along with the Streamlit framework for building the user interface.

The code begins by importing the required libraries and defining a background style for the Streamlit application. It then sets up the user interface, allowing the user to enter multiple URLs from which they want to extract product information. The number of URLs is determined by the user using a number input widget.

For each URL entered, the code defines a function named get\_product\_info() that sends an HTTP GET request to the URL, retrieves the webpage content, and parses it using BeautifulSoup. It extracts various product details such as the name, category, star rating, ratings, reviews, cost, seller information, and other sellers if available. The extracted information is stored in a dictionary. The code then loops through the entered URLs, calls the get\_product\_info() function for each URL, and appends the extracted product information to a list called data. After extracting the product information, the code checks if any data was obtained. If there is data, it creates a pandas DataFrame using the data list and displays it using Streamlit's dataframe function. The DataFrame is then converted to a CSV format, which can be downloaded by the user using a download button.

Finally, the code includes an else statement that displays a message if no data is available. The Streamlit framework allows the application to be launched in a web browser, providing an interactive interface for entering URLs, displaying the extracted data, and enabling the download of the data as a CSV file.

* **Step-by-Step Process**

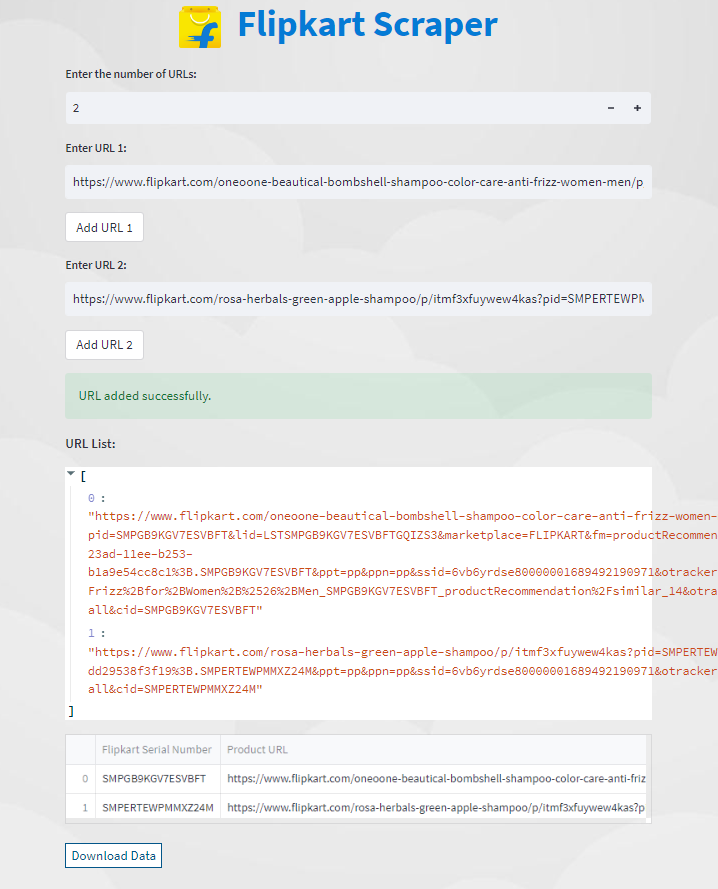
1. Importing Libraries: The necessary libraries are imported, including base64, requests, BeautifulSoup from the bs4 module, pandas for data manipulation, and streamlit for building the user interface.
2. Setting the Page Background: The page\_bg variable contains CSS code that sets the background image for the Streamlit application. It uses the background-image property to specify the URL of the image and the background-size property to control its appearance.
3. Setting the Application Header: The code uses the markdown function from Streamlit to display an application header. It includes an image of the Flipkart logo and sets the title using HTML tags. The unsafe\_allow\_html=True parameter is added to allow the rendering of HTML content in Streamlit.
4. Initialize Session State: The session state feature of Streamlit is utilized to store and manage the entered URLs. The code checks if the urls variable exists in the session state. If not, it creates an empty list.
5. Inputting URLs: A loop is created based on the num\_urls variable, which represents the number of URLs to be entered. For each URL, a text input field is provided for the user to enter the URL. If the user clicks the corresponding "Add URL" button, the URL is added to the urls list in the session state, if it is not already present. Appropriate success or warning messages are displayed based on the input validity.
6. Displaying URL List: The entered URLs are displayed using the st.write() function. This provides a visual representation of the URLs that have been entered.
7. Defining the get\_product\_info() Function: This function takes a URL as input and performs web scraping on the Flipkart product page. It sends an HTTP GET request to the URL, retrieves the webpage content, and parses it using BeautifulSoup. The function then extracts various product details such as the product name, category, star rating, ratings count, reviews, cost, seller information, and other sellers. The extracted information is stored in a dictionary and returned as the output.
8. Extracting Product Information: The code loops through the entered URLs stored in the session state. For each URL, it calls the get\_product\_info() function to extract the product information. The extracted information is appended to the data list.
9. Displaying Extracted Data: If data is available in the data list, it is converted into a pandas DataFrame using pd.DataFrame(). The DataFrame is displayed using the st.dataframe() function, which renders it as an interactive table in the Streamlit application. Additionally, the code modifies the "Seller" column by removing the last three characters to enhance the presentation.
10. Downloading Data: If there is data available, the code converts the DataFrame into a CSV format using the to\_csv() method. It then encodes the CSV data using Base64 encoding and generates a download link using an HTML button. The link allows the user to download the data as a CSV file.
11. Displaying No Data Message: If no data is available, the code uses the st.markdown() function to display a custom HTML message indicating that no data is present.

* **Snap Shots of the Application**

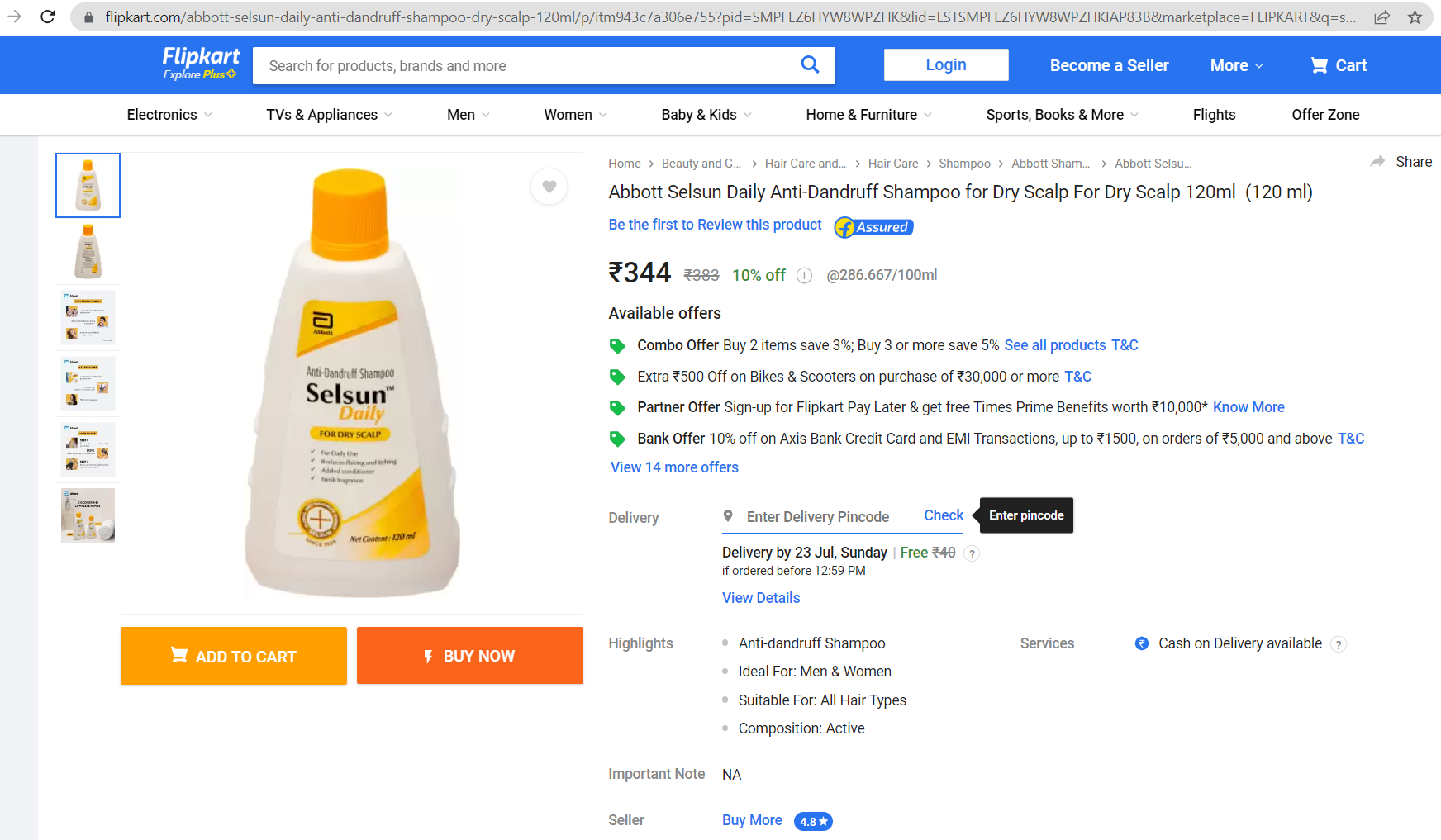
A screenshot of a computer

Description automatically generated

Web Application



Web Application with data entered.

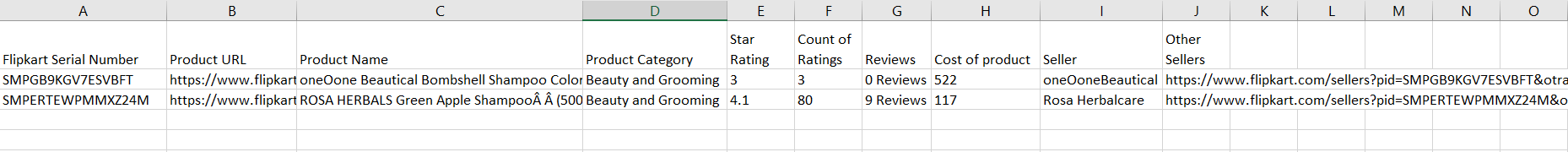


Example of product whose data will be extracted

A screenshot of a computer

Description automatically generated

Pop-up whenever a link is added to the web application



Output obtained in the Excel Sheet

**Code:**

A screenshot of a computer code

Description automatically generated

A computer code with text

Description automatically generated

A screen shot of a computer code

Description automatically generated

A screen shot of a computer code

Description automatically generated

A white screen with black text

Description automatically generated

A computer code on a white background

Description automatically generated

**Pharma Sales Dashboard**

The dataset is built from the initial dataset consisted of 600000 transactional data collected in 6 years (period 2014-2019), indicating date and time of sale, pharmaceutical drug brand name and sold quantity, exported from Point-of-Sale system in the individual pharmacy. Selected group of drugs from the dataset (57 drugs) is classified to the following 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
* Sales data are resampled to the hourly, daily, weekly and monthly periods. Data is already pre-processed, where processing included outlier detection and treatment and missing data imputation.

**Dashboard**

A screenshot of a computer

Description automatically generated

The data has been pre-processed, including outlier detection and treatment, as well as missing data imputation. The pre-processing step ensures the data quality and prepares it for analysis.

The created dashboard provides visual representations of the sales data using different types of charts. The pie chart compares the sales of drugs in the R03 (drugs for obstructive airway diseases) and R06 (antihistamines for systemic use) categories. This comparison gives an overview of the relative sales contribution of these two therapeutic classes.

The donut chart compares the sales of all six drug categories (M01AB, M01AE, N02BA, N02BE/B, N05B, and N05C) in the dataset. This chart helps in understanding the distribution of sales across these categories and identifying any dominant categories.

The area chart displays the sales of M01AB and M01AE drugs over a period of 12 months. This chart allows for a visual comparison of the sales trends of these two categories within a specific timeframe.

The stacked bar chart compares the sales of N02BA, N02BE/B, N05B, and N05C drugs, showcasing the relative contribution of each category to the total sales. This chart provides insights into the sales distribution among different psycholeptic and analgesic drug categories.

The dashboard includes slicers for selecting the month, year, and quarter, enabling users to explore the sales data for specific time periods.

Overall, the dashboard and its visualizations provide a comprehensive overview of the sales data for different drug categories, allowing for easy comparison and analysis of sales trends over time.

**Score Card Automation SAP**

In the Score Card Automation SAP project, one of the areas I focused on was a window where medicines could be added or removed based on categories. During my analysis, I discovered a bug related to a specific category called HEPC. The bug allowed duplicate entries to be added for this category, which resulted in data redundancy. This issue needed to be addressed because duplicate entries can lead to inaccuracies and inconsistencies in the system. By identifying and reporting this bug, I helped ensure data integrity and improve the overall functionality of the SAP system.

Another aspect of the Score Card Automation SAP project involved calculating scores for salespeople based on a predefined calibration matrix. This score determined the incentives they would receive. The calibration matrix values were initially created in an Excel sheet and then uploaded into the SAP system. During my work, I identified a bug in the calibration matrix where the score generation for salespeople who achieved between 61 to 74 points was not properly calibrated. The issue stemmed from incorrect data being injected into the calibration matrix. This bug affected the accuracy of the scores and had implications for the incentives received by salespeople. By identifying this bug, I helped ensure that the calibration matrix was corrected, resulting in accurate score calculations and fair distribution of incentives to the sales team.

Identifying and addressing these bugs in the Score Card Automation SAP project was essential for maintaining data accuracy, eliminating redundancies, and ensuring fairness in the scoring and incentive system. By resolving these issues, the overall functionality and reliability of the SAP system were improved, providing a more efficient and effective platform for evaluating sales performance and motivating the sales team.

**KEY LEARNINGS**

The key learnings from the Flipkart Web Scraping project include:

* Understanding of Web Scraping Techniques: The project provided an opportunity to gain a deeper understanding of web scraping techniques and their application in real-world scenarios. By working on the project, you acquired knowledge of libraries like Beautiful Soup and learned how to navigate and extract relevant information from web pages. Understanding the fundamentals of web scraping opens up possibilities for automating data collection and analysis tasks.
* Automation and Efficiency: By automating the data collection process, you were able to significantly reduce manual effort and save time. This experience showcased the power of utilizing programming languages like Python to automate repetitive tasks. The project demonstrated how automation can improve efficiency and productivity by freeing up valuable human resources for more complex and strategic tasks.
* Problem-Solving and Troubleshooting: Web scraping projects often involve encountering challenges and obstacles during the process. Throughout the project, you likely faced issues such as handling dynamic web pages, dealing with inconsistent data structures, or addressing anti-scraping measures. These challenges provided an opportunity to develop problem-solving and troubleshooting skills, enabling you to find creative solutions and workarounds.

The key learnings from the Pharma Sales Dashboard project include:

* Data Visualization for Insights: The project provided an opportunity to understand the importance of data visualization in extracting insights from complex pharmaceutical sales data. By creating visualizations such as pie charts, donut charts, area charts, and stacked bar charts, you gained insights into sales trends, comparisons between drug categories, and other key metrics. This experience highlighted the effectiveness of visual representations in communicating data-driven insights to stakeholders.
* Utilizing Power BI for Data Analysis: The project involved working with Power BI, a powerful data visualization and analytics tool. By leveraging Power BI's features and functionalities, you gained hands-on experience in creating interactive dashboards, exploring data, and generating meaningful visualizations. This experience enhanced your proficiency in using Power BI for data analysis and decision-making in the pharmaceutical industry.
* Data-driven Decision Making: The Pharma Sales Dashboard project emphasized the importance of data-driven decision-making in the pharmaceutical industry. By analyzing and visualizing sales data, you were able to identify patterns, trends, and outliers that informed decision-making processes. This experience highlighted the significance of leveraging data to drive strategic initiatives, optimize sales strategies, and improve overall business performance.

The key learnings from the Score Card Automation SAP project include:

* Importance of Thorough Testing: The project highlights the significance of comprehensive testing to identify and address bugs or issues in the SAP system. Through diligent testing, I was able to discover and report bugs related to data redundancy and incorrect calibration matrix values. This underscores the importance of thoroughly testing all aspects of the system to ensure its reliability and accuracy.
* Attention to Detail: The project emphasized the need for meticulous attention to detail when working with complex systems like SAP. By carefully reviewing and analyzing the system's functionalities, I was able to identify discrepancies and inconsistencies that could have a significant impact on data integrity and the accuracy of calculated scores. Paying close attention to details is crucial to ensuring the smooth operation and reliability of such systems.
* Data Integrity and Accuracy: The project underscored the criticality of maintaining data integrity and accuracy within the SAP system. Bugs related to data redundancy and incorrect calibration matrix values could lead to incorrect scoring and incentives for salespeople. Ensuring data accuracy and integrity is crucial in decision-making processes, performance evaluation, and incentive programs. It emphasizes the need for robust data validation and quality control measures.

**CONCLUSION**

In conclusion, my internship experience at Abbott as a Business Intelligence Intern has been incredibly rewarding and insightful. Throughout the internship, I had the opportunity to work on various projects and technologies such as Python, Power BI, Advanced Excel, and SAP. These experiences allowed me to apply my skills and knowledge to real-world business challenges in the pharmaceutical industry.

I would like to express my sincere gratitude to Abbott for providing me with this invaluable internship opportunity. The guidance, support, and mentorship I received from the team were instrumental in my professional growth and development. I am grateful for the trust placed in me and for the chance to contribute to the company's goals and objectives. I look forward to applying the knowledge and skills gained during my internship in future endeavors. The experience at Abbott has strengthened my passion for business intelligence and has further solidified my commitment to pursuing a successful career in this field.

Once again, I extend my deepest appreciation to Abbott for the incredible internship experience, and I am confident that the skills and experiences gained during this period will have a lasting impact on my professional journey.

**REFERENCES**