A PROJECT REPORT

ON

SALES ANALYSIS AND PREDICTION USING PYTHON

BY

HARIHARAN R

REG.NO. -202300690

(2023 – 2025)

Submitted in partial fulfilment of the requirements for qualifying

POST GRADUATE CERTIFICATION IN BUSINESS ANALYTICS

SYMBIOSIS CENTRE FOR DISTANCE LEARNING

DECLARATION

This is to declare that I have carried out this project work myself in part fulfilment of the Post Graduate Certification in Business Analytics Program of SCDL.

The work is original, has not been copied from anywhere else and has not been submitted to any other University/Institute for an award of any degree/diploma.

Name: Hariharan R

Signature: Hariharan R

Place: Chennai

Date: 12-05-2024

ABSTRACT

One description of business intelligence is a tool that assists associations in making informed opinions about their internal operations. Effective data association and conversion into information, which forms the base of knowledge in decision- timber, may be achieved with its help. The establishment may significantly contribute to timely business performance monitoring and prompt response to the external business terrain with the aid of the applicable business technology result. This design looks at data and uses the Python programming language to fantasize it in order to develop deals output that shows yearly deals success. Using multiple direct retrogression in Python, we've bettered the data output with advertising data and generated an interactive dashboard that shows the estimated deals grounded on investment inputs from colourful marketing channels.

INDEX

Sr.No.	Contents	Page No.
1	Introduction	5
2	Analysis of Work Done and Design	6
3	Learning experiences on Business / Technology	10
4	Conclusion	12
5	Bibliography / References	13

1. INTRODUCTION

Concerning data, ordinary activities of a company lead to huge amounts of data generation. Corporations find it hard to understand and utilize bulky data sets that are randomly collected over time. Again, quick and appropriate decisions must be made for an organization to respond fast enough to changes in its external environment. Therefore, business intelligence becomes imperative in mining for information from the data collected as well as converting them into knowledge for decision making purposes. Certain business units may be largely dependent on IT assistance due to the simplification of smart business software. For example, rather of asking their technical colleagues to produce data, employees may create reports that are easier to examine in order to generally monitor the success of sales in the sales department. In order to increase value for sales performance, it can expedite tasks and boost productivity. To make sure operational requirements are explicit, changes are received and quick decision making is enhanced; it is important to study how businesses operate. This way, one will be able to join IT and business plans, validate the necessary steps and provide benefits for organizations as well as their clients. The capability of Business Analysis allows data driven decision making as well as more simplified technical communication. The subjects crucial in guiding specific changes within an organization are tactics and strategy in business. Business analysis can make companies become driven by analytics thus helping them to discover new ways of making decisions basing on information through data. It is important for deeper insights into enterprises to be provided so as to foster corporate transformation towards being customer-centric innovative operation. Building an outcome report based on sales and utilizing Python to design a sales dashboard that reflects sales performance will investigate how a dataset may empower businesses and offer interactive visualizations and business intelligence capabilities.

2. ANALYSIS OF WORK DONE AND DESIGN

Figure 1. Report creation process (Methodology) using Python

This methodology's steps, as seen in the block diagram above, are defined in the following.

1. To start, gather all the databases and duplicate them. Then, clean and modify the dataset to improve its usability.

2. When developing reports, dig down on personnel and products using filters.

3. Calculate revenue from sales by unit sales, revenue per product, and revenue per unit sold.

4. Determine the income increase on a daily basis and make a drill-down table with pertinent columns.

2.1 Extracting the Data

Data about products, and sales are included in the dataset for sales analysis. Kaggle was the source of the datasets used to forecast sales.

2.2 Transforming the Data

A machine learning method has to be used in order to anticipate sales, even though a sales analysis database may be easily imported. Data must thus be created before being utilized in machine learning models, meaning that it cannot be used in its original state. Thus, it is necessary to modify the data before loading it. The knowledge extractor uses this approach to tackle problems that it is currently unaware of. Pre-processing requires properly cleaned and prepared data.

This will eventually be used in the formula, where anticipated sales are calculated as the total of several of those channels plus the coefficient.

$$y = b0 + b1*x1 + \quad + bn*xn$$

When the channels are represented by x1,...,xn, the expected sales by y, and the coefficient is denoted by b. Python was utilized to separate those coefficients and calculate the expected sales value through the application of linear regression.

2.3 Algorithm used

Among machine learning algorithms, linear regression is the most popular and extensively utilized. The goal or dependent variable and the response or independent variables are related linearly through its utilization. Based on the following equation, the linear regression model is created:

$$yb0 + b1x1 + b2x2 + b3x3 + \ldots\ldots + bnxn$$

Where x1, x2, x3, and y are regarded as the target and intercept variables respectively whereas b1, b2, b3,...,bn represent corresponding coefficients and xn stand for independent variables.

This approach seeks to identify the line of greatest fit between the target variable and the data's independent variables. This is accomplished by identifying the leading ideal values for each b. To have the best fit, the projected value must have the least amount of error and be extremely near to the real values. The distance between the data points and the fitted regression line is known as the error, and it is often computed using the following formula:

Error = y1 - y2

where y2 is the expected value and y1 is the actual value.

2.4 Loading the Data

Product, and sales data were all included in the product sales database that Google Colaboratory imported for the sales analysis dashboard. And when the advertising database

was transformed, a database containing a variety of data was placed into the sales forecast outcome.

2.5 Report Creation using Python

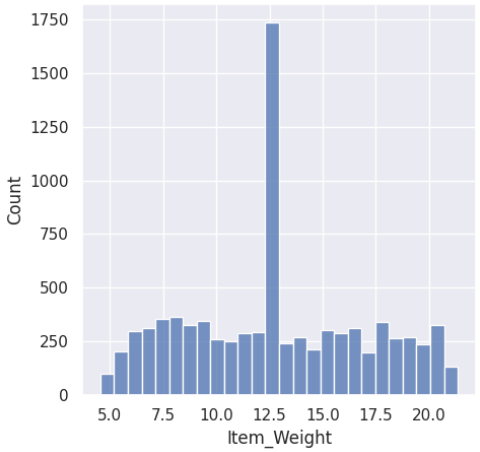

Figure 2. Item's Weight Distribution

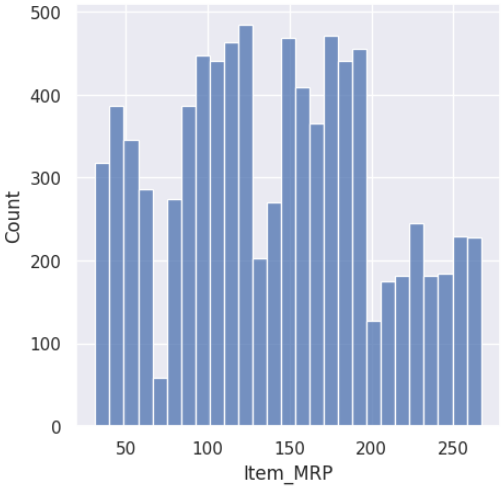

Figure 3. Item's MRP wise Distribution

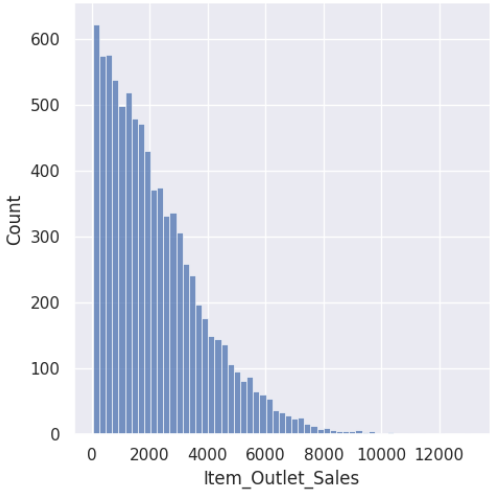

Figure 4. Outlet Sales Distribution info.

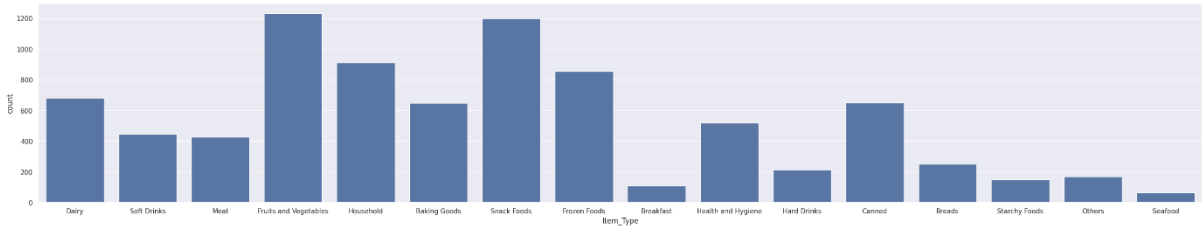

Figure 5. Item Type Distribution

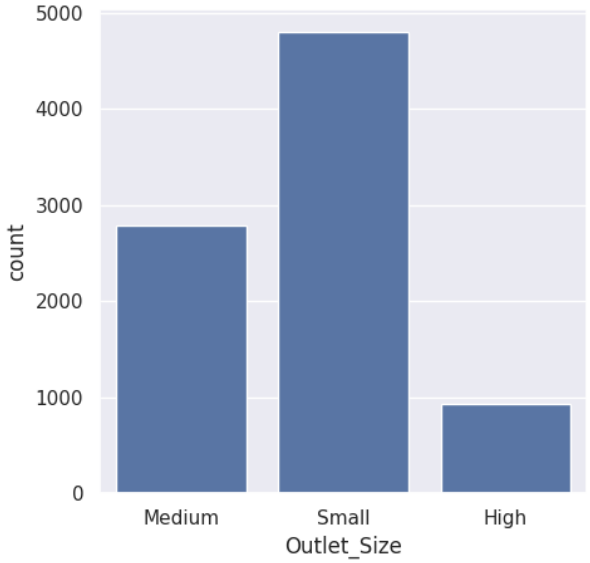

Figure 6. Outlet Size Distribution

2.6 Result

The aim of this project was to identify the best relationship between the dependent and the independent variables in the dataset that the multiple linear regression method was amenable with; that aim was accomplished by the researcher. For statistical visualization in Python, seaborn is the best library, and for data manipulation, pandas are a perfect library to use. It was made possible to estimate and predict the sales making it simpler to analyse and forecast pattern of sales.

3. LEARNING EXPERIENCES ON BUSINESS / TECHNOLOGY

Introduction

In today's data-driven environment, companies are constantly generating more and more data. However, using this information to make informed decisions remains challenging. Achieving the company's performance requires business intelligence in order to draw meaningful conclusions from data. This chapter discusses how lessons learned and challenges faced are combined with sales analytics and forecasting projects using machine learning and Python programming.

Lessons Learned

The course of the project brought forward a few important lessons learned: first and foremost, there was the importance of data quality and pre-processing. Gathering and cleaning of datasets turned out to be critical steps in ensuring the accuracy and reliability of the analysis. Moreover, understanding the business context and domain knowledge has been instrumental in helping to interpret the results accordingly. Effective communication and collaboration between the business and technical teams were of utmost importance for the success of the project.

Experiences Imbibed

The importance of the project was a positive experience in combining theoretical knowledge with practice. The second one was understanding the necessity of parallel development of technical solutions along with the business idea. The analysing of data from the actual calculation sheets helped to explore the machine learning algorithms better, most notably linear regression, and how they work for the prediction of sales. Other impacts include improvements in problem-solving and creative leverage to solve arising issues met during the research.

Bridging Theory and Practical Practice

the project was an intermediate step and a bridge between theoretical ideas and their implementation in practice. The fact is that the courses on data analysis and machine learning provided with the necessary theoretical background. All those learned concepts then had to be used on real data sets to apply the knowledge in practice. In the case of the project, linear regression algorithms created a model of the dependence of leading sales indicators on various other indicators of customer behaviour. Based on these patterns, it was possible to predict sales indicators. Moreover, the opportunity to directly apply the skills I received in courses enhanced my understanding of how the algorithms work and how the data processing process is built.

Conclusion

In conclusion, the sales analysis and prediction project has allowed me to acquire crucial skills in both the business and technology industries. It underlined the role played by data-driven decisions, good communication, and cooperation. Through this project, the theory and practice were to be bridged, thus improving skills in data analysis, machine learning and problem-solving. Looking ahead, lessons from the current experiences as well as the strength received shall be taken to the next levels. These in turn will be the source of power to sustain the growth in business analytics.

4. CONCLUSION

The following conclusions may be drawn from the results discussion and the multiple linear regression application using Python:

1. The top of the list is the step, where the data is being cleaned.

2. Based on this case study, the following suggestions are made for ongoing improvement:

● The pre-built database has to be updated in order for it to meet the information demands of several companies that the corporation owns and uses in an effective and efficient manner.

3. Firms may use predictive analytics technology to identify and foresee upcoming events and formulate and implement timely decisions that create competitive advantage in the industry and result in higher sales revenue.

4. As of now, the best model would be discerned through evaluation and comparing the prediction models with their metrics like accuracy. The multiple linear regression model has been found to be the most suitable model for this purpose after the evaluation phase with the help of different channels.

Finally, the results of the predictive analytics work will be to spot the new clients and it will be queer the sales forecast and sales letter mailing campaigns which drive high sales in return.

The sales research and forecast project is part of a learning curve that has proven to be worthwhile and has resulted in a number of valuable lessons for a business and for technology. The campaign acted as a reminder of the importance of teamwork, effective communication, and data-based decision-making. I also benefited from analysing real data sets with machine learning and data analysis, which broadened my problem-solving skills.

This experience will inevitably push the boundaries for greatness of business analytics for the future. I possess the skills and knowledge that enable me to face new problems and to come up with new ideas. Now I am convinced that I come up with input which guarantees the growth of the business via my good knowledge of data analysis and applying it in practical conditions.

5. BIBLIOGRAPHY / REFERENCES

[1] Research Paper- Building A Sales Dashboard for A Sales Department by Using Power BI by Lan Huang Lan Huang

[2] "Data Strategy: How to Profit from a world of Big Data, Analytics and The Internet of things" by Bernard Marr

[3] "Big Data Demystified: How to use Big Data, Data Science and AI to Make Better Business Decisions and Gain Competitive Advantage" by David Stephenson PhD

[4] "The Development of Performance Dashboard Visualization with Power Bi as Platform" by Surlisa Widjaja, Tuga Mauritius

[5] "Descriptive and Predictive Analytics on Adventure Works Cycle: A Corporate Decision Making" by Yew Liong Lim, Raheem Mafas

[6] Delen, D., Sharda, R. & Turban, E. 2014. Business Intelligence: A Managerial Perspective on Analytics. 3rd edition. England: Pearson Education Limited.

[7] PT. Indomobil Sukses Internasional, "Annual Report 2017 PT Indomobil Sukses Internasional Tbk .," www.indomobil.com, vol. 2017, 2017.

[8] L. Moss and S. Atre, Business Intelligence Roadmap. 2003.

[9] D. Delen and H. Demirkan, "Data, information and analytics as services," Decis. Support Syst., vol. 55, no. 1, pp. 359–363, 2013.

[10] D. Appelbaum, A. Kogan, M. Vasarhelyi, and Z. Yan, "Impact of business analytics and enterprise systems on managerial accounting," Int. J. Account. Inf. Syst., vol. 25, no. March, pp. 29–44, 2017.

[11] Braun, V. & Clarke, V. 2006. Using Thematic Analysis in Psychology. Article from ResearchGate. Accessed 21 January 2019.

[12] https://www.researchgate.net/publication/235356393_Using_Thematic_Analysis_in_Psychology

[13] Brinkmann, S. 2013. Qualitative Interviewing. Book from ProQuest. Accessed 20 January 2019. https://ebookcentral.proquest.com/lib/laurea/detail.action?docID=1274289

[14] Castleberry, A. & Nolen, A. 2018. Thematic analysis of qualitative research data: Is it as easy as it sounds? Article from ScienceDirect. Accessed January 2019.

[15] https://www.sciencedirect.com/science/article/pii/S1877129717300606

[16] Clark, D. 2017. Beginning Power BI: A Practical Guide to Self-Service Data Analytics with Excel 2016 and Power BI Desktop. Book from ProQuest. Accessed 2 January 2019. https://ebookcentral.proquest.com/lib/laurea/detail.action?docID=4815394&query=clark

[17] Mckay, R. 2018. Data Visualization Example: A Look into Modern Visual Innovation. Posted 3 October. Accessed 26 December 2018. https://www.tapclicks.com/innovative-data-visualization-examples/

[18] Microsoft. No date. Building Blocks of Power BI. Accessed 20

December 2018. https://docs.microsoft.com/en-us/learn/modules/get-started-with-power-bi/3-buildingblocks-of- power-bi

[19] Microsoft. 2016. Overview of Service Manager OLAP cubes for advanced analytics. Accessed 3 January 2019. https://docs.microsoft.com/en-us/system-center/scsm/olap-cubes-overview?view=sc-sm1807

[20] Microsoft. 2018. Intro to the dashboard for power BI designers. Accessed 5 January 2018. https://docs.microsoft.com/en-us/power-bi/service-dashboards

[21] M. Mitri, "Teaching Tip Active Learning via a Sample Database: The Case of Microsoft's Adventure Works," Journal of Information Systems Education, vol. 26(3), 2015.

[22] "A Guide to the Business Analysis Body of Knowledge: Draft Material for Review and Feedback,"
International Institute of Business Analysis, 2006

[23] Knaflic C., "Storytelling With Data: A Data Visualization Guide for Business Professionals", Wiley, 2015.

[24] D. M. Velumani, "A Study on Consumer Buying Behaviour towards Nokia Mobile in Erode District," IOSR Journal of Business and Management (IOSR-JBM), Vol. 16, No. 12, pp. 52-54, 2014.

[25] D. L. I. A. M. Asoc. Prof. Elīna Gaile-Sarkane, "Consumer behavior-changing: methods of evaluation,"
Trends of Economics and Management, Vol. 3, No.4, 2013

[26] Shafique U., Qaiser H., "A Comparative Study of Data Mining Process Models (KDD, CRISP-DM, and SEMMA)," International Journal of Innovation and Scientific Research, Vol. 12, No. 1, pp. 217-222, 2014.

[27] J. H. Rüdiger Wirth, "CRISP-DM: Towards a Standard Process Model for Data Mining", IFIP International Conference on Digital Forensics, 2017

[28] Vente J, Waal A.D, Willers C., "Specializing CRISP-DM for Evidence Mining," SpringerLink.

[29] M. Faber, "Visualizing Student Research though a Multi-Conditional Branching - Choose Your Adventure-Style Tutorial," Semantic Scholar, 2015. Accessed on Jun 19, 2019. [Online]. Available:
https://www.semanticscholar.org//paper/Visualizing-Student-Research-through-a-Branching-%22- Faber/10684e7187e06ccadb11e185ab 1d58c8cf90a35e