School of Art and Creative Technologies

University of Greater Manchester

Evaluation of Room and Desk Booking Trends for Strategic Recommendations in NHS Facilities.

Anthony Ogunna

Dr. Pradeep Hewage

Dr. Anchal Garg

A Research Paper Submitted in the Partial Fulfilment for

The Award of MSc degree in Data Analytics and Technologies

SEPTEMBER 2025.

University of Greater Manchester, Deane Road, Bolton. BL3 5AB

DECLARATION

This is to certify that this thesis titled "Evaluation of Room and Desk Booking Trends for Strategic Recommendations in NHS Facilities" is entirely original with no submissions to other educational institutions and the sources utilised in this study have been properly cited and referenced.

ACKNOWLEDGEMENT

I begin by saying thank you with all my heart to God Almighty for His guidance, wisdom, and grace throughout the study of my MSc and completion of this dissertation. I would like to express my deepest gratitude to my research supervisor, Dr. Pradeep Hewage, for his expert advice, constructive feedback, and continued motivation. I further appreciate immensely Dr. Anchal Garg, who managed this project under the industry partnership, for her useful inputs and motivation. I wish to thank particularly Maywoods Limited, NHS consultants, for providing access to the booking datasets used as the foundation for this study on the optimisation of space utilisation for NHS facilities. Through this, they facilitated this study. Finally, I offer heartfelt appreciation to my mother, Mrs. Anthonia Ogunna, for being my best inspiration, through prayers, faith, and unwavering encouragement. I also wish to thank my family and friends for their cooperation and understanding during this exercise. This dissertation would not have been possible without the help and contributions of all those involved.

ABSTRACT

This project examines the NHS booking system with the general aim of identifying redundant bookable units and proposing the minimum number of units to facilitate successful booking service delivery. Twenty three (23) datasets were merged via SQL following recreation of the system's relational flowchart, yielding an initial setup of five (5) districts, twenty three (23) buildings, and four hundred and three (403) bookable units. Following preprocessing and data cleaning, exploratory data analysis was conducted with python to analyse booking behavior and usability effectiveness. The result showed that among the years under review (2016 - 2026), 2024 had the highest bookings, while the 2025 and 2026 bookings were incomplete as they had partial-year capture and future bookings in process. Tuesday and 9:00 a.m. were found to be the busiest booking times, and one-hour and eight-hour time slots dominated the usage. Most strikingly, over 80% of the units showed poor negative correlations between bookings and usage, reinforcing the point that usage evaluation alone is not sufficient as an index of efficiency. For instance, Unit 42 had typical high utilisation but was in operation in 2016, 2017 and 2018 only. The audit revealed that a combined total of one (1) district, seven (7) buildings, and one hundred and twenty three (123) units were redundant, reducing the system to four (4) districts, sixteen (16) buildings, and two hundred and eighty (280) bookable units. This minimum capacity is brought forth as the minimum needed sustainable capacity that can maintain NHS operations without undercutting delivery of booking services. Five predictive models were attempted and tested, among which Random Forest was found to be the most accurate, but prediction was only used as an aid and not the focus of the project. It demonstrates that there are substantial optimisation opportunities for NHS booking capacity, which generate data-driven recommendations such as retiring redundant bookable units and reforming utilisation measures for enhanced efficiency without reducing resilience in services.

TABLE OF CONTENTS

WORD COUNT : 17,823

TABLE OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

NHS	National Health Service
MAE	Mean Absolute Error
SQL	Structured Query Language

CHAPTER ONE

INTRODUCTION

Optimal use of space and resources is an imperative element of service provision in the National Health Service (NHS). Desks and meeting rooms are regularly booked in most of the facilities, but all the bookable units available are not made optimal use of. It results in wasted capacity, higher costs, and difficulties in managing demand. The primary goal of this study is to explore booking and usage patterns in the NHS system, identify redundant bookable units, and suggest the minimum number of configuration needed to offer services effectively. This was done by the integration of various districts' and buildings' datasets and processing. The study applies data cleansing, exploratory data analysis, and predictive modelling to identify booking patterns, peak times, redundant units, efficiency rates and forecast. This enables the study to provide evidence-based, action-oriented recommendations that will help in optimising the maximum use of NHS bookings, improving efficiency, and making sure services are robust without losing out on availability.

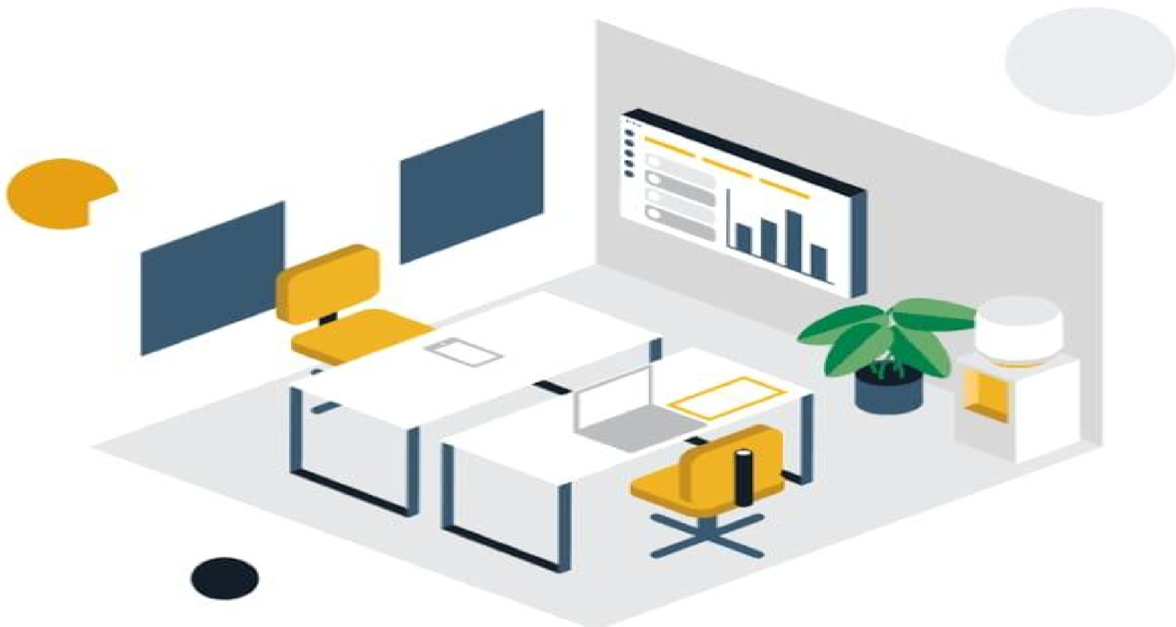


Fig. 1: Meeting room (Workero, 2025).

1.1 Background of Study

Across many large organisations, the need to align physical workspace with evolving work practices has come more and more into focus. In healthcare settings, this matter is critically pertinent in view of the contribution of workspace arrangement to staff satisfaction and organisational output (Harris, 2015). Misaligned or insufficient space provision can lead to inefficient use and workflow of the facilities available (Bodin and Theorell, 2019). But informal observations shows that there is a significant number of these bookable spaces that are likely to be underused because they no longer accommodate modern working practices or because there has been a decline in on-site presence (Water et al., 2018). The greater utilisation of hybrid and remote work, especially since the COVID-19 pandemic, has further disconnected workspace supply from real-time demand (Shetty et al., 2024). Consequently, NHS estate managers are relying on data-based approaches to determine which physical sites are required and which can be optimised or re-directed without impacting service provision (Cawood et al., 2016). Evidence across other domains has indicated that data analysis, in this case through evaluation of booking trends within records, can reveal patterns of use and aid the detection of underutilised assets (Fleming et al., 2012). Where implemented in healthcare environments, such methodologies not only enable more intelligent decision making but also aid in the transition towards optimised, adaptive estates that better reflect changing demands of public service delivery (Tabak, 2009). An evidence-based understanding of true utilisation behaviour can form the foundation for well informed decisions regarding retention, consolidation, reuse, or release of surplus space. The present research investigates how quantitative analysis can illuminate such behaviour and thus contribute to strategic estate planning across NHS sites.

Booking optimisation is important in NHS facilities because meeting rooms and desks are essential for staff collaboration, training and administration. When these bookable units are not efficiently managed, it can lead to redundancy, wasted capacity or shortage during peak demand. This can reduce staff productivity and affect the quality of healthcare service delivery. By identifying redundant bookable units and making sure the right number of resources are available, booking optimisation helps the NHS minimise running cost, improve booking service efficiency and promote effective staff operations.

1.2 Problem Statement

Many NHS buildings possess desks and meeting rooms that are constantly under-occupied, resulting in operational inefficiencies as well as wasteful costs. Existing practices to determine what assets are underutilised and redundant often rely on informal observation or local judgement , with managers lacking a system-wide, fact based framework to guide estate optimisation. Without robust measures of utilisation, the NHS may continue to spend on or refurbish space that is of negligible practical benefit to staff. There is thus a need to scan empirical booking data, to quantify the extent and nature of underuse and redundancy, and to translate those findings into effective estate plans.

1.3 Research Questions

This study endeavour to evaluate how NHS buildings, specifically. meeting rooms and desks are utilised today in different districts and offer practical advice. To serve as the starting point for the analysis and to meet the objectives of the research, the following research questions have been established:

1) What are the existing yearly booking patterns for the bookable units in the NHS buildings under review?

2) At which (a) days of the week and (b) hours of the day does booking activity reach its peak?

3) Based on the booking occasion history, which usage duration recorded the highest number of bookings?

4) What is the correlation between total yearly booking and utilisation rate of the bookable units?

5) How many bookable units, buildings, and districts can be considered redundant within the years under review up to 2025, based on booking occasion data?

6) What is the minimum number of buildings and bookable units currently required to sustain NHS booking system operations in the districts under review without negatively impacting service delivery?

7) (a) Which predictive model provides the best fit for forecasting the total yearly bookings of bookable units for 2026? ?

(b) What is the graphical comparison of the predicted values of best fit model with the actual values?

(c) What is the forecast total yearly booking for bookable units in 2026?

1.4 Research Contribution to Knowledge

This study contributes to knowledge by offering data driven, replicable two-dimensional approach (temporal and spatial evaluation) for measuring space utilisation across public sector property portfolios, in the context of the NHS. While prior literature has recognised the importance of good workplace layout and space utilisation in the healthcare setting, much of that literature is devoid of empirical examination based on operational booking data. This study fills that gap by applying quantitative methods to actual data sets, offering objective assessment of desk and meeting room use in various NHS buildings and districts. The research adds to the body of work by Introducing a methodology that integrates SQL database management, python for exploratory data analysis and machine learning for predictive modeling to demonstrate how information from booking systems in multiple facilities can be collected, examined, and interpreted to identify booking and usage patterns with rich visualisations. It also prescribes spatial and temporal limits for redundancy detection, a technique that remains poorly developed in the existing literature on NHS estate planning. By integrating facility management and data analytics, this work advances existing operational research models within healthcare and includes a scalable model that can be extended for use in other public service organisations with comparable issues in estate optimisation. Besides, the study contributes new information by revealing some of the underlying patterns of behaviour and organisational use which are not easily accessible using traditional estate valuations. It provides evidence that can inform strategic decision-making and policy formulation on space consolidation, re-use and planning for service delivery. Through this, the research bridges the gap in research-based theory on space management and adoption of practical data driven approach in large, resource constrained organisations like the NHS.

1.5 Aim and Objectives of the Research

1.5.1 Aim

The aim of this research is to leverage on data analytics to quantify and optimise desk and meeting room utilisation across NHS buildings by identifying redundant resources and recommending minimum booking assets suitable to effectively run the booking system without negatively impacting on service delivery. Being able to see how physical space is really being utilised according to actual booking data can serve to inform strategic decisions in estate planning so that space is used to good effect and remains able to contribute to the achievement of operational objectives.

1.5.2 Objectives

In order to fulfill this purpose, the research establishes the following specific objectives:

I) Carry out in-depth analysis of literature to identify gaps in existing literature in space utilisation.

ii) To use data-driven methods to explore and identify temporal and usage patterns.

iii) To examine desk and meeting room booking data in selected NHS buildings in order to evaluate actual space usage rates and booking trends across the facilities during the years under review.

iv) To identify redundant bookable units that are feasible for consolidation, relocation, removal from booking system or functional redesign using exploratory data analysis

v) To build a predictive model to help forecast future bookings.

vi) To facilitate data-driven, evidence-based recommendations for NHS managers to permit more informed and strategic estate decisions.

By fulfilling the above objectives, the study seeks to bring together and develop the evidence base for NHS workspace planning and additionally provide a contribution of an operational model for public sector space optimisation.

1.6 Scope of this Study

The study focuses on understanding booking patterns and optimising usage of NHS bookable units. The analysis of data was conducted on the basis of booking records from 2016 up to 2026 (being future bookings), providing a rich longitudinal view of trends of the use of space. The geographical region covers NHS estates dispersed over five districts, across series of administrative blocks. The focus is exactly on bookable rooms such as desks and meeting rooms, reserved as such in the NHS electronic booking systems.

The investigation integrates various datasets booking history, bookable unit properties, area designs (spaces, buildings, districts), team and service affinities, and booking metadata like type, status, and cancellation. SQL and python based tools were applied in analysis and processing with the aim of producing evidence-based outcomes to support NHS managers in space optimisation decision making and strategic estate planning.

1.7 Thesis Organisation

The study is divided into five chapters.

Chapter 1 contains the introduction to the study, background, problem setting context, aim and objectives, research questions, and contribution.

Chapter 2 consolidates literature concerning workspace usage, facility management analysis, and post pandemic workplace practice, referring to theoretical foundations and research requirements.

Chapter 3 discusses methodology, detailing the datasets, data merging processes, analytical techniques and ethical considerations.

Chapter 4 presents and discuss the results and findings. It also includes predictive models evaluation and answers the research questions.

Chapter 5 conclude the thesis, providing recommendations for consolidation or reconfiguration of space, outlining study limitations, and suggesting directions for future research.

The conversation as a whole shows how insight built on data can support evidence-based estate choices, and hence allow the NHS to justify physical workspace to meet shifting organisational needs.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Traditionally, space usage within organisations such as the NHS has been managed on an ad hoc basis, based on staff reports, feedback, and observational research. Booking logs, interval surveys, and physical room and desk audits were employed by facility managers in order to arrive at a guess of usage rates. These were intended to capture a snapshot of demand and identify areas that were underutilised but were carried out at intervals, meaning decisions were taken on the basis of snapshots and not on rolling data.

While these traditional approaches provided some information, they were incomplete and misleading. Observational audits and manual surveys are time-consuming, labour-intensive, and prone to error. They are unable to detect up-to-the-minute fluctuations in demand, and staff opinion-based usage often results in partial or incomplete information. In large resource-constrained institutions like the NHS, these gaps can lead to wasteful utilisation of desks and rooms, overestimation of off-peak times, and unnecessary maintenance of redundant facilities.

A data-driven approach takes advantage of the use of computer analysis and web-based reservation systems to observe what is really going on with space utilisation and desk usage. Instead of making an educated estimate or conducting surveys, it observes patterns of use and history of bookings to know which rooms are reserved and which aren't. This allows organisations to spot capacity waste, create better plans, and match resources with reality. For the NHS, where budgets are often tight, this is more than worthwhile because it saves time, saves money, and allows staff to become more effective. In short, it turns guess work into fact and allows managers to make more informed decisions in the future.

2.2 Prisma diagram

Below is the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram showing the literature review flow chart;

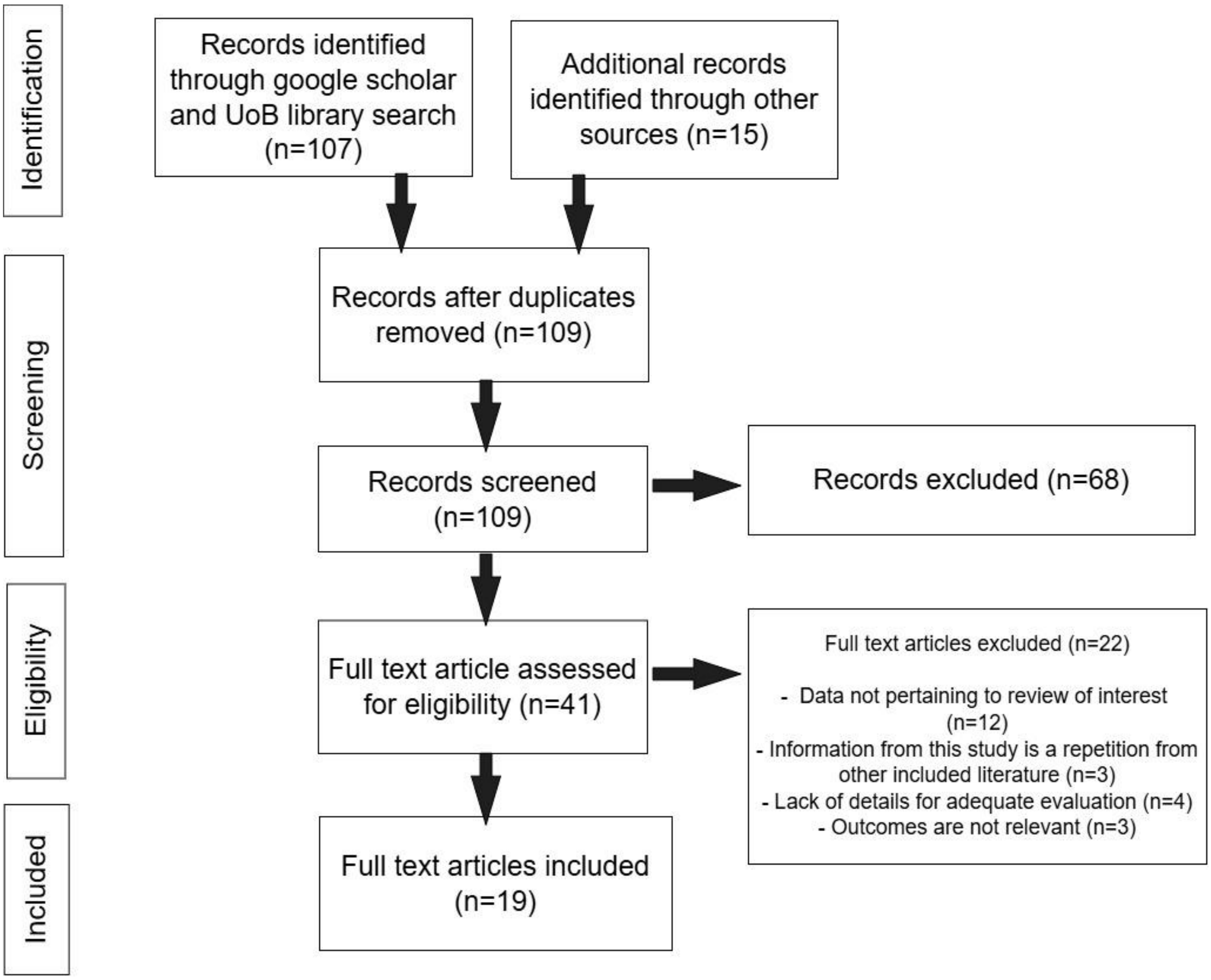


Fig. 2 : Prisma diagram

2.3 Review of some Literature

Effective management of the physical space of the healthcare organisational space is an old and yet more urgent problem. In such publicly supported health organisations as the National Health Service (NHS), whose operating pressures and resource limitations are ever-present, optimum use of meeting rooms and desk space is at once an operational concern. Over the past twenty years, more focus has been laid on linking space utilisation and design to organisational objectives, well-being of the workforce, and outcomes of service delivery (Cawood et al., 2016; Kim et al., 2016; Shetty et al., 2024). This research situates itself within this context and with a perspective of using data analysis for the investigation of patterns in NHS center booking behavior and utilisation and ultimately arrive at evidence-based recommendations for better space management.

Within healthcare specific environments, Lange, Häne and Windlinger (2022) evaluated the potential for space efficiency, presenting a helpful example of how models based on information can highlight potential for better resource optimisation. Their work underpins research such as this one, which seeks to go beyond description and towards prescriptive conclusions based on quantitative analysis. In fact, the use of simulation and optimisation methods to resource allocation also became more widely accepted in operational environments. Ghasemkhani, Torabi and Hamid (2023) proposed a hybrid simulation and optimisation approach to operating room capacity planning, recognising the potential of strong modelling to support decision making in the face of uncertainty. They write here with particular reference to surgical capacity, but their methodological observation is how far such analytical techniques may be applied to workspace utilisation problems.

Rapport et al. (2019) further discussed the impact of workplace layout in facilitating organisational productivity and resilience. These researches point out that the usage of space must be achieved not just in a quantitative manner but as a matter of relational and cultural terms.

Conceptually, Harris (2015) explained how the changing dynamics of the workplace, be it through the movement towards flexible and hybrid arrangements requires updating traditional estate management models. This emphasis is repeated in more contemporary studies on activity based workplaces. To show, Appel-Meulenbroek, Groenen and Janssen (2021) explored how users perceive activity based offices, emphasising how perceived productivity is intertwined with user satisfaction and workspace design. Hoendervanger et al. (2019) also considered the influence of use flexibility on switching behaviour and users' satisfaction. These findings have extremely high applicability in the NHS context, where hybrid working cultures and evolving attitudes towards space are becoming ever more common.

Voordt and Jensen (2023) integrated evidence that healthy workplaces ensure employee satisfaction, productivity, and cost savings, a reminder that space optimisation is as much about value creation as cost containment.

The environmental team interface has also been a focus of attention in recent studies. Peavey and Cai (2020) have developed a systems model linking environmental factors to clinical teamwork and noting that workspace layout can facilitate or complicate collaborative working. In a world where multidisciplinary working increasingly becomes unavoidable in healthcare delivery, such information is doubly important in facilitating space design for integrated care.

The variation in methodologies among the studies, from systematic reviews and simulation modeling, qualitative ethnographies, and smart technology solutions spans the subtlety of recognising and optimising space utilisation. While various studies have been carried out in the corporate or academic environments (Fleming et al., 2012; Tagliaro et al., 2021), its application in healthcare estates is relatively in its early stages. Tabak (2009) built user simulation systems for office space utilisation, laying the foundation for even more sophisticated occupancy modelling presented only in recent years for the hospital environment.

There is still progress to be made though amidst this increasing evidence base. For the NHS, empirical research is needed to make the most of detailed booking data and apply rich analytics to deliver actionable recommendations. While perception surveys and questionnaires have addressed key subjective dimensions of workspace experience (Kim et al., 2016; Bodin Danielsson and Theorell, 2019), few have formally tested time-stamped usage records for a series of facilities and organisational units. This is a serious gap, in an age of increased usage of computerised reservation systems and a risk that it might be used to make more informed estate management decisions.

The study therefore seeks to bridge this divide through application of data analysis to booking records in NHS in an attempt to quantify usage patterns and make areas of possible improvement clear. Adopting the methodological advancements outlined in literature i.e., intelligent monitoring (Tagliaro et al., 2021), simulation and optimisation (Ghasemkhani et al., 2023), and performance indicator frameworks ,the present research recommends building an evidenced base applicable to inform space consolidation, repurposing, and redesign strategic planning. In particular, it also alludes to the fundamental role of organisational culture and user experience in defining the occupation of space, as highlighted by qualitative office and healthcare space studies in turn (Water et al., 2018; Rapport et al., 2019).

Lastly, empirical evidence from the literature is conclusively that space use is a multi-determined dynamic process controlled by organisational policy, user behaviour, technology, and design. With the sophisticated measurement and modelling practice of magnitude, utilising the facilities in healthcare estates management, particularly that of the NHS, is useful research evidence. This study has employed the underlying research knowledge of experts across disciplines to develop a data-informed model for informing and maximising the use of space with the ultimate goal of enabling quality, efficient, and effective delivery of health care.

2.4 Related Works

On a scale of 1 to 5, where 5 represents the highest level of relevance, Table …. below lists publications in this field of study and highlights their significance to this study

Publications	Relevance	Methodology	Technique	Strengths	Limitations	Knowledge contribution
Sutton, L., Tarrant, C., Willars, J., Coats, T., Simmonds, M., Mclean, D., Boyle, A.,	4	Qualitative interview and observation-based study.	Operational management practice thematic analysis.	Single focus on NHS-specific operating environments. It illustrates how groups work in co-use environments.	Quantitative measurement of space use or bookings does not occur. Primarily focused on fulfilling use rather than total	Shows how NHS staff cope with resilience and workflow, and indirectly how space planning affects service delivery.

Dreesbeimdiek, K., Richter, S., Oyedijo, A. and Roland, D. (2025)					workspace use.	
Shetty, R.S., Kamath, G.B., Rodrigues, L.L., Nandineni, R.D. and Shetty, S.R. (2024)	4	Systematic review of evidence on physical environment effect.	Evidence synthesis and thematic categorisation.	Very comprehensive summary of a series of studies. It is a very current publication showcasing contemporary evidence.	Not greatly interested in quantitative booking figures. Perhaps more focused than the specific topic of desk/room	Highly strong synthesis of knowledge concerning how physical spaces affect staff and why space optimisation is

					utilisation.	needed.
Peavey, E. and Cai, H. (2020)	4	Empirical literature systematic review	Evidence synthesis, thematic categorisation	1) Economically synthesises disparate studies to form in-depth framework. It also recruits clinical environments, in keeping with NHS setting.	Unearths teamwork beyond quantitative booking patterns. It does not provide primary utilisation statistics.	Provides a conceptual framework linking environment design and teamwork success in support of evidence-based workspace programs.

Tagliaro, C., Zhou, Y. and Hua, Y. (2021)	5	Conceptual and literature review of sensor-based and IoT-based measurement methods.	Descriptive analysis of the application of smart technologies to monitor occupancy.	Up to date discussion of digital measurement methods. It Illustrates use of data analytics to increase precision of utilisation.	More focus on promise of technology rather than empirical data. Also, commercial application examples rather than health-service delivery.	Does offer new evidence on integration of smart monitoring with facility management, transferrable to NHS estate optimisation with immediate impact.
Voordt, T.V.D. and Jensen, P.A. (2023)	4	Empirical literature review and case study synthesis and review.	Comparative analysis, thematic synthesis.	Evidence review of benefits emerging from healthy workplace design. It	Potentially more focused on general offices than clinical or administrative	Enhanced the case for investment in space planning on the argument that

				considers both financial and operational perspectives.	NHS environments. It has preference for health outcomes over booking utilisation.	healthy workplaces optimised generate a range of organisational dividends.
Ghasemkhani, A., Torabi, S.A. and Hamid, M. (2023)	4	Hybrid modelling with discrete-event simulation and mathematical optimisation.	Simulation-based optimisation for operating room capacity planning.	Emphasises sophisticated modelling under uncertainty. It declares how simulation enhances resource allocation.	Applies specifically to operating room scheduling, not general workspace use. It will need adaptation for NHS desk/meeting room allocation.	Demonstrates how simulation and optimisation combined can improve capacity planning, with applicability to space demand modelling

						methodology.
Aw, S.F. and Rohayu, O. (2020)	2	Cross-sectional survey of tertiary hospitals with surveys and observational methodologies.	Quantitative data collection and descriptive statistics for environmental factors on staff.	Measures direct staff perception of conditions within the workplace. It plays understanding of physical space effect on worker satisfaction within context.	None of the context measures space booking practices nor contain data on time or space utilisation.	Adds insight into the nature of the quality of the workplace in healthcare environments and their impact on staff wellbeing, providing background information as to

						why space optimisation is valuable to health outcomes.
Bodin Danielsson, C. and Theorell, T. (2019)	4	Survey survey of office workers, contrasting perceptions by gender and office layout.	Quantitative statistical analysis of questionnaire responses.	Extremely strong emphasis on user experience and satisfaction. It delivers clear indication of the correlation between workplace	Limited to general office settings, not healthcare settings. It does not contain booking system information or objective measures	Relates to the value of aligning workspace layout with staff needs and preferences, something that is worth remembering

				configuration and perceived contribution to work.	of usage.	in the course of NHS space optimisation planning.
Cawood, T., Saunders, E., Drennan, C., Cross, N., Nicholl, D., Kenny, A., Meates, D., and Laing, R. (2016)	5	Human-centred design approach with stakeholder engagement, environmental survey, and recursive design.	Mixed-methods, qualitative staff feedback augmented by design evaluation.	Good participatory approach with end users coupled with simple transferral to healthcare workspace design.	No quantitative booking measures. Also, excess design over empirical measurement of utilisation.	Shows how interactive, user-driven design processes can create spaces that more readily facilitate their healthcare staff and operational performance, an

						absolute gem for data analysis.
Harris, J. (2019	4	Conceptual and applied discussion with corporate real estate strategy case studies exemplars.	Descriptive analysis of actual-workplace change programs.	Real-world demonstration of value of data-driven space planning. It identifies links between utilisation intelligence and organisational	Corporate environments but not public healthcare. It moderately lower emphasis on statistical methods.	Strengthening argument for use of space utilisation analytics in strategic planning and workforce priority alignment.

				strategy.		
Harris, R. (2015	4	Literature review and conceptual analysis of trends in office space utilisation.	Integration of practice and research themes within the industry.	Context for why the utilisation of space is a rising top issue. It connects future of office space with flexible work.	Not empirical, not on booking systems or healthcare. No original data, no statistical analysis.	Offers conceptual model of drivers for change of workspace, a good context to put NHS study in broader trends.

Kim, J., Candido, C., Thomas, L. and De Dear, R. (2016)	4	Employee survey of staff in non-territorial offices.	Quantitative analysis of self-reported job satisfaction, perceived productivity, and health measures.	Empirical findings on the human impact of shared workspace environments. It offers insight into compromises of desk-sharing policies.	Context primarily commercial office settings. It is based on perception rather than actual utilisation data.	Explains how desk allocation models can influence staff experience, a relevant consideration when analyzing NHS desk booking behaviour.
Lange, S., Häne, E. and Windlinger, L. (2022)	5	Empirical research integrating usage measurement with spatial analysis in health facilities.	Space occupancy and efficiency measurement; quantitative analysis.	Hospital-specific nature. It integrates spatial measures of efficiency closely with operational issues.	Potentially specific to UK hospital systems of one kind. It does not mention desk booking	Strong evidence on identification of space efficiency opportunities and action in hospitals, in

					systems at all.	your case for justification of your NHS research.
Nejati, A., Shepley, M., Rodiek, S., Lee, C. and Varni, J. (2016)	3	Multi-method study with survey, observation, and spatial analysis.	Quantitative and qualitative analysis of break room design features.	Places emphasis on the environment in enabling staff wellbeing. It is methodologically sound.	Generalisability only to meeting room or workstation booking behavior. Break rooms were the focus, not working space	Demonstrates supporting evidence that workspace arrangement, unbooked space as well, has a significant influence on staff experience, including satisfaction.

Rapport, F., Auton, E., Cartmill, J., Braithwaite, J., Shih, P., Hogden, A. and Clay-Williams, R. (2019	4	Multimethod qualitative study protocol integrating ethnographic observation and interviews.	Qualitative methods - observation, interviews, and document analysis.	Rich qualitative findings regarding how hospital work environments construct resilience and productivity. Also provides insight into actual staff experience.	No quantitative data on use. It makes use of protocol paper instead of finished findings.	Stresses the importance of workspace design and layout in operational resilience and how they supplement quantitative space usage evidence in your study.

Sadatsafavi, H., Walewski, J. and Shepley, M.M. (2015)	3	Perception measurement by healthcare staff using surveys.	Quantification of ratings and perceptions for different spaces.	Inform staff about workspace issues. Good foundation for interpreting space sufficiency.	More perception based than usage behavior. It does not capture booking or temporal usage data.	Contributes to knowledge of how healthcare personnel evaluate workspaces, in the process aiding indirectly the argument for evidence-based optimisation.
Water, T., Wrapson, J., Reay, S. and Ford, K. (2018)	4	Qualitative study of spatial practice by observation.	Ethnographic approaches and qualitative analysis.	Rich understanding of spatial use in practice. It is healthcare-specific focus.	Paediatrics-oriented rather than administrative offices with no	Emphasises the ways in which workers travel and take up space and implies

					measurable measure of utilisation.	that material spaces need to be mapped against work practice.

Table 1 : Related works summary

2.5 Summary of recent techniques and technologies on space utilisation

Recent studies on the use of space have shown a shift away from the traditional method of observation towards more advanced and technology-driven approaches. Internet of Things (IoT) and sensor monitoring are now being used to provide real-time usage information, and this is promising greater precision and consistency in measuring the use of space (Tagliaro et al., 2021). Concurrently, optimisation and simulation models, such as discrete-event simulation with mathematical optimisation, have been applied to develop capacity planning in conditions of uncertainty, and there is clear scope for application within NHS booking systems (Ghasemkhani et al., 2023). Spatial analysis with usage measurement has been applied to advance efficiency opportunity identification in health clinical environments (Lange et al., 2022). In addition, more participatory approaches such as human-centred design, where stakeholders are implicated in the process of experiencing design a workspace, provide another perspective of how enhancing utilisation can be done (Cawood et al., 2016). They are used to complement one another to demonstrate growing emphasis on digital, analytical, and participatory approaches as better ways of optimising space with lessons that can be shared in NHS building management.

2.6 Summary of research gaps

Limited quantitative measurement on space usage

Much of the current research employs staff opinion, questionnaireing, or qualitative observation rather than proper examination of booking and utilisation data, which would indicate limited quantitative proof of what is actually utilised in NHS or healthcare-specific environments.

Healthcare focus but not on meeting/desk booking

Whereas some research explores hospitals or theatres, little has been touched on in administrative and meeting rooms essential to the functioning of NHS staff, whereas most of the evidence comes in clinical rooms or general offices rather than NHS booking systems.

Disconnect between theory and practice

Review and conceptual articles do often suggest the potential of intelligent monitoring and data-driven planning but rarely apply such strategies on real NHS datasets, with a void in demonstrating how theories of space optimisation can be applied in practical capacity planning.

Insufficiency of predictive, data-driven modelling for NHS booking systems

Advanced modelling studies have a specific interest in clinical scheduling such as operating theatres, yet leave room for the application of historical booking history data, exploratory analysis, and forecast modelling for NHS meeting and desk space.

This study closes the research gaps by shifting the focus from theory and perception to a methodical data-driven analysis of actual NHS booking patterns and configuration, identifying redundant bookable units and providing actionable recommendations for staff meeting space and desk space resource optimisation.

CHAPTER 3

METHODOLOGY

3.1 Overview of the Methodology Chapter

The methodology chapter of this research provides a comprehensive description of the philosophical perspective, research design, data management strategies, and analysis methods implemented in studying space utilisation in NHS estates. As public sector organisations increasingly experience the issues of operational effectiveness and estate efficiency, the ability to quantify precisely and predict the utilisation of workspace such as desks and meeting rooms becomes an important aspect of strategic planning. This chapter outlines the systematic and rigorous approach employed in this study to derive actionable insights from historical booking data with the long term objective of applying evidence based and data driven analysis to inform estate management decisions.

The chapter is organised in the following way: the chapter begins with an examination of the underlying research philosophy of the study, then the approaches, outlining the preprocessing and exploratory processes. The chapter continues with discussion of the general research methodology and design and a flow diagram to give a visual representation of the procedural steps. Subsequent sections closely track the illustration of each action performed in the analysis, from the accumulation of data to the development of descriptive and predictive insight. A preliminary section also discuss the evaluation of predictive models and application of the best fit model in forecasting bookings for 2026. As much as this model constitutes an ancillary part of the study, it shows how modeling can be incorporated in spatial planning schemes. The chapter concludes with justification for each methodological decision to ensure that it aligns with the research aim and objectives.

3.2 Research Philosophy

Research philosophy is an initial decision in planning any research study, as it decides the researcher's epistemological and ontological stance, and thus affects the methodological framework and analytic procedures employed. Three dominant research philosophy paradigms, positivism, interpretivism, and critical realism have distinct perceptions of reality and knowledge construction process.

Positivism believes that social facts can be studied objectively, like natural sciences, based on experience and rationality from the assumption of social reality as measurable and external to human consciousness. Based on hypothesis testing, formal methods, and quantitative analysis of data, positivist research aims for generalisable trends from empirical observation and rationality. The paradigm is objectivist, reliable, and replicable and is most appropriate for research studies trying to determine cause and effect or trends within sizable sets of data.

Interpretivism, however, has a constructivist approach to reality, arguing that reality is not a given fact but constructed from people's knowledge and social environments. Researchers in this paradigm seek to learn about meanings, experiences, and interactions from the participants' perspective, most often through qualitative methods such as interviews, focus groups, or ethnography.

Critical realism offers a mature middle ground, acknowledging that there is an objective world but not acknowledging that our knowledge of it is constructed through social forms, cultural practice, and historical context. Critical realism combines qualitative and quantitative methods of analysing complex social phenomena in detail.

As a result of the focus of this research, i.e., in examining data from the NHS booking systems to understand usage patterns and make decisions regarding estate planning, a positivist paradigm has been used. This is a suitable approach with the type of data, empirical, quantitative, and amenable to statistical and computational analysis and the need to deliver objective, generalisable, and usable outcomes with which to inform operational planning. As compared to interpretivist approaches, that would call for consideration of the unique user experience, or critical realism, which often includes the addition of more pervasive structural analysis, positivism is best suited to the need for systematic, evidence-based knowledge from measurable and observable events. In addition, the emphasis on methodological rigor, replicability, and validity under positivism is also most suitable for practical use of results in a healthcare management setting so that counsel is well founded on high quality empirical facts.

3.3 Research Methodology

This research adopts a data-driven, quantitative method of research with the ultimate goal of extracting useful patterns from historical booking records for NHS estates space utilisation to identify redundant booking resources. The space utilisation analysis follows a two-dimensional approach that focus on the usage and booking count analysis for holistic evaluation. Methodology is comprised of several connected steps: data cleaning and preparation, exploratory and descriptive analysis, spatial-temporal pattern identification, and utilisation classification. Apart from descriptive statistics and data visualisation, modeling is employed to illustrate how predictive analytics could enhance static dashboards and reports.

The underlying analysis process is founded on the paradigms of exploratory data analysis (EDA) and spatial-temporal analysis, both facilitating the identification of usage patterns across time and location. Cutoff for underutilisation is determined as relative and absolute standards, for example, average occupancy rate and inactivity period. For the purpose of this study, the NHS standard utilisation cutoff is 50% for usage evaluation. This will act as a reference in the process of classifying. While machine learning techniques are not the focus of this project, including modeling is justified as a means of demonstrating the scope for forward-planning space tools. This is of particular interest to NHS managers wishing to transition from reactive towards more proactive management of the estate.

As described in chapter two, these are the research gaps to be addressed in this study:

I. Limited quantitative measurement on space usage

ii. Healthcare focus but not on meeting/desk booking

Iii. Disconnect between theory and practice

iv. Insufficiency of predictive, data-driven modelling for NHS booking systems.

To address these gaps, the following methods are employed;

3.3.1 SQL for Merging Data

The merging process followed a systematic approach to integrate twenty three booking datasets to form the consolidated NHS booking dataset for this study as seen in the merging flowchart shown in fig. 3.1 below; Once all datasets containing the respective booking data have been merged into the Final merged dataset using a sequence of SQL operations as seen in Appendix A, a deduplication process is then carried out to delete duplicate rows. This provided data consistency and avoided the presence of the same bookings in the analysis in multiple forms. The consolidated booking dataset is the end result of this pipeline of integrated data, and it is the main analytical booking dataset for this research called Maywood merged dataset.

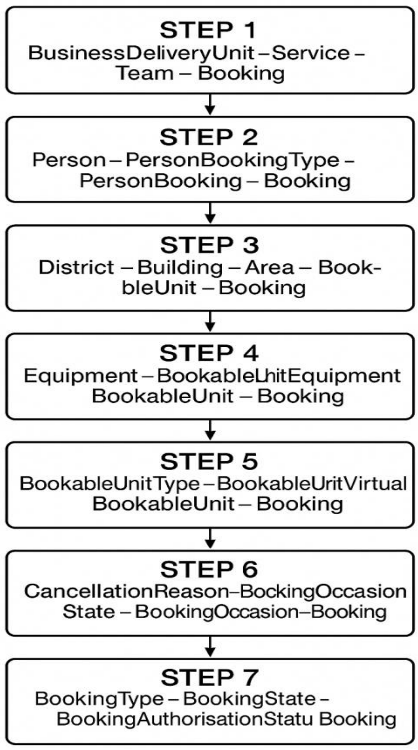


Fig. 3: Merging flowchart

3.3.2 Python for Preprocessing and Data Analysis

Python was used for data preprocessing and exploratory data analysis of the merged dataset. Elimination of inconsistencies, missing value handling, cleaning data, outliers detection among others were done during preprocessing. Feature engineering was also carried out to create new variables such as yearly booking count, actual booking hours and actual building operating hours to calculate utilisation rate. Exploratory data analysis (EDA) came later to find patterns in booking, peak booking day and time, identify redundant bookable units, correlation between usage and booking count and so on. Python was used because of its rich set of libraries that enable data manipulation, statistical analysis and visualization to be done efficiently.

3.3.3 Machine Learning for Predictive Modelling

Five predictive models were selected, namely: Bayesian Model, Linear Regression, Random Forest, Gradient Boosting Regressor, and Linear Support Vector Regression (SVR). These were selected to represent both linear approaches (Bayesian and Linear Regression) and non-linear approaches (Random Forest, Gradient Boosting, and SVR). The choice of selection is to provide room for comparison between baseline models and advanced ensemble approaches that can handle non-linear booking patterns. The model performance evaluation is done using Mean Absolute Error (MAE), which is a measure that defines the average difference between actual and predicted values. MAE was chosen for simplicity of interpretation, The evaluation process involved training and testing the models using historical booking information and comparing values of MAE to determine the model that achieved the highest predictive accuracy.

3.4 Methodology Chart Flow

Below is the methodology flowchart for this study as shown in fig. 3;

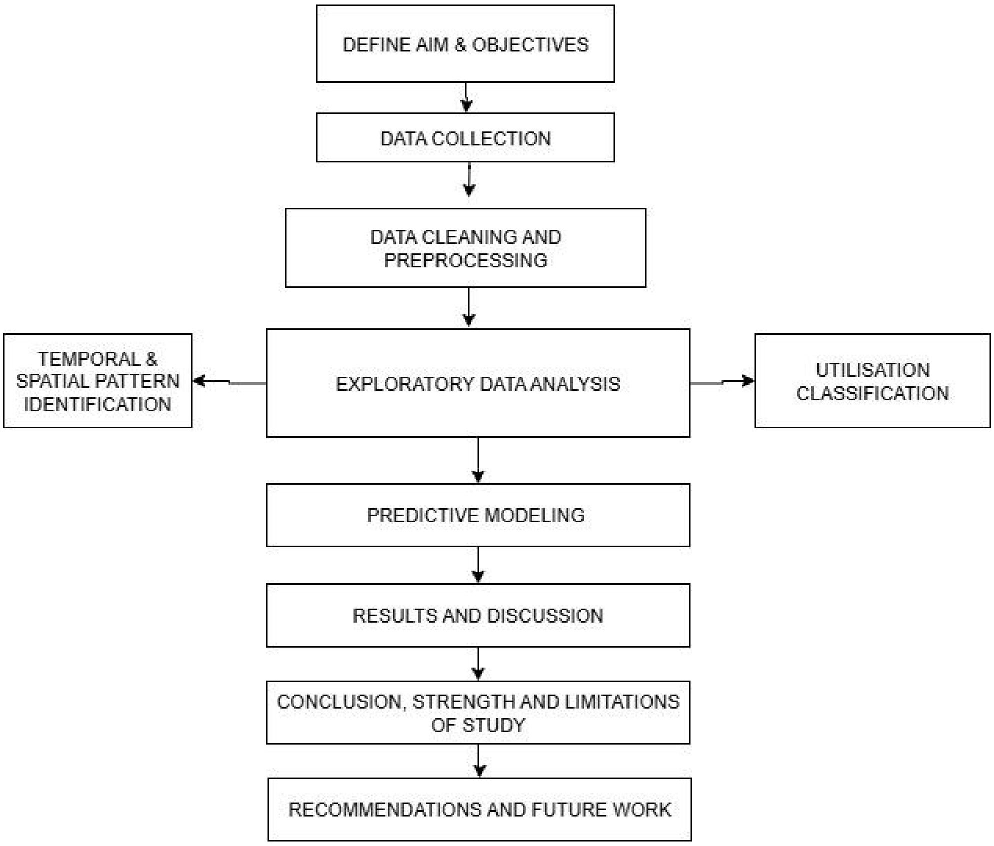


Fig. 4: Methodology flowchart

3.5 Research Design

The study employed a quantitative, descriptive, exploratory, and predictive research design, based on secondary data analysis of NHS booking system records. There were no interventions and controlled experiments thus carried out; instead, the study entails observing, measuring, and interpreting naturally occurring patterns of room booking behavior.

The design is predominantly cross-sectional in nature, the purpose of which is to take a snapshot of booking and usage over a designated temporal time period. It also contains a longitudinal analysis component insofar as it is drawn from timestamped data and thus can be employed to examine trends over time, e.g., year-to-year fluctuation in patterns of booking. The design also contains a component of comparison and thus it is possible to measure differences in levels of utilisations between various organisational units, buildings, and geographic locations in the NHS trust.

This multi-part assignment enables the following amount of strategic questions, among others:

Are there some rooms or buildings that are chronically redundant and underused?

Do booking levels for the rooms vary significantly by day of the week or time of day?

Using this study design, the research aims to gain a general understanding of prevalent space utilisation trends within the NHS and where there is high operating inefficiency and identify where resource optimisation and strategic reprofiling can occur.

3.5.1 Steps for data analysis

The analysis was implemented through a systematic process that consisted of seven major steps.

Step 1: Understanding the Problem

The fundamental issue for this research is reducing the space assigned versus space utilised gap. Underutilisation is a source of wastage and potential loss for increased use of space. It is important to estimate the size of this problem and explore ways in which optimal space utilisation may be attained.

Step 2: Aggregating the Data

Data were pulled from reservation systems to be utilised on multiple NHS sites. The systems track such data as space reservations, name of person who reserved, date of function, number of people attending, and facility data. The data collection had a number of NHS sites and a number of facility types, providing some sense of space behaviour usage.

Step 3: Data Cleaning and Preparation

Prior to proceeding to analysis, the data is cleaned for consistency. Duplicate records were removed, and standardised date and time fields into a uniform format. Missing values were replaced with reasonable estimates or excluded where it would not have had any impact on conclusions in general.

Step 4: Exploratory Data Analysis (EDA)

The data is then explored in order to learn more about its shape and to expose preliminary trends and patterns.

Step 5: Visualising the Data

Visualisations proved helpful in enabling good understand of the data. Heatmaps showed holistic correlation across attributes. Bar charts gave direct measurement levels. Boxplots helped to identify any possible outlier and line plot gave a simplified trend flow of temporal evaluation.

Step 6: Predictive Modelling

Five models were selected and their mean absolute error evaluated to identify the best fit model for prediction.

Step 7: Drawing Strategic Insights

Lastly, findings are translated into actionable recommendations for NHS estate managers. These are recommendations to streamline redundant and under-occupied bookable units, and encourage more sustainable booking practices through intervention.

Separating the analysis in this organised way is aimed to provide NHS estate managers with actionable, evidence-based recommendations that can optimise use of space and business efficiency.

To do all of these as discussed above, data is needed. Below is a description of the dataset;

3.5.2 Description of the Dataset

The dataset used in this study was extracted from NHS booking systems and provides a rich tapestry of data on the usage of desks and meeting rooms across an extremely diverse range of buildings and districts.

The data were captured in twenty three (23) individual datasets as listed below:

Area.csv

AvailabilitySearch.csv

BookableUnit.csv

BookableUnitEquipment.csv

BookableUnitType.csv (x2)

BookableUnitVirtual.csv

Booking.csv

BookingAuthorisationStatus.csv

BookingOccasion.csv

BookingOccasionState.csv

BookingState.csv

BookingType.csv

Building.csv

BuildingOpeningHours.csv

BusinessDeliveryUnit.csv

CancellationReason.csv

District.csv

Equipment.csv

person.csv

PersonBooking.csv

PersonBookingType.csv

Service.csv

Team.csv

which were subsequently merged into a single dataset using SQL for the purpose of analysis. The aggregated dataset that is the dataset for this project has 43 variables and 1048576 entries, and includes relevant dimensions such as timing information (booking start and end times), space identifiers (bookable unit titles, building Ids, district names, etc), user information (team and service information), and metadata about bookings (booking status, authorisation levels, number of visitorsetc). Below are the descriptions of the dataset attributes grouped into six (6) categories;

Booking Details

Booking_Id : A unique identifier that is assigned to each individual record of a booking.

Booking_MeetingTitle : The designation or title assigned to the meeting or event associated with the booking.

Booking_BookingTypeId : A category code to be used to indicate the type of booking (i.e., regular, recurring, one-off).

BookingType_Title : The explanatory title that is associated with the type of the booking.

Booking_BookingStateId : Identifier to display the current state of the booking (e.g., active, cancelled, completed).

BookingState_Title : Booking state description.

Booking_BookingAuthorisationStatusId : Authorisation code indicating whether the booking is authorised or not.

BookingAuthorisationStatus_Title : Human-readable authorisation status representation (e.g., approved, pending, rejected).

Organisational Hierarchy

Booking_TeamId : Reference linking the booking to the organisational team making the booking.

Team_Title : Name of the team making the booking.

Service_Id : Identifier for the service the booking relates to.

Service_Title : Service title or name by which the booking has been placed.

BusinessDeliveryUnit_Id : Business unit ID through whom the service is being delivered.

BusinessDeliveryUnit_Title : Title of business delivery unit.

Personnel and Attendance

PersonBooking_Id : Person-level booking record ID unique.

PersonBooking_BookingId : Directs the person booking towards the main booking ID.

Person_Id : Unique ID of the person creating or attending the booking.

PersonBookingType_Id / PersonBookingType_Title : Identifies the type or role of person in the booking (e.g., attendee, organiser).

Bookable Unit (Space) Details

BookableUnit_Id : Unique identifier for a bookable space (e.g., room, desk).

BookableUnit_Title : Title or number of the bookable space.

Area_Id / Area_Title : Identifier and title of the area of a building where the bookable unit is located.

Building_Id / Building_Title : Building identifier and title where the bookable unit is situated.

District_Id / District_Title : Administrative district identifier and title where the building is situated.

EquipmentList : List of booked or reserved equipment within the bookable unit.

BookableUnitType_Title : Bookable unit type (e.g., meeting room, hot desk).

BookableUnitVirtual_ChildBookableUnitId : For virtual units, defines related physical units.

BookableUnitVirtual_Capacity : Maximum occupancy capacity of the bookable unit.

Booking Occasion (Temporal Data)

BookingOccasion_Id : Unique ID per occurrence of a booking occasion.

BookingOccasion_BookingId : Refers the occasion to the parent booking.

BookingOccasion_StartDateTime / BookingOccasion_EndDateTime : Date/time stamps representing the scheduled start and end dates/times of the booking.

BookingOccasion_DAY : Day of week on which the booking occurred.

BookingOccasion_NumberOfVisitors : Number of individuals who were intended to attend the booking.

BookingOccasion_CancellationReasonId : Key for the reason for cancellation (if applicable).

CancellationReason_ReasonText : Textual description of the reason for cancellation.

BookingOccasion_BookingOccasionStateId / BookingOccasionState_Title : Status of a booking occasion (e.g., active, cancelled).

Building Operational Hours

BuildingOpeningHours_DayId : Identifier for the day of the week.

BuildingOpeningHours_IsOpen : Flag to indicate whether or not the building is open on that day.

BuildingOpeningHours_OpeningTime / BuildingOpeningHours_ClosingTime : Scheduled opening and closing times for the building on some days.

At a first look, the dataset contained some issues common in big administrative datasets, including duplicate columns, sparse attributes, and semantically redundant fields. The process of data auditing and cleaning was an exhaustive one with the objective of enhancing analytical efficiency and relevance. The columns below were removed from subsequent analysis due to reasons:

Booking_Id:1, Booking_Id:2: These are duplicate references to the parent booking identifier and are not needed.

PersonBooking_BookingId, BookingOccasion_BookingId: These are foreign key associations already represented in existing booking fields and do not offer any further analytical value.

BookableUnit_Id:1, BookableUnit_Id:2: These are duplicate references to the parent BookableUnit_Id field.

BuildingOpeningHours_ columns*: Day ID inhabitants, opening and closing times, open/closed status, these columns are chiefly empty and non-representative of actual booking behavior, better captured through timestamps in the booking occasions.

CancellationReason_ReasonText, BookingOccasion_CancellationReasonId: These possess high levels of missing values and are just pertinent in cancellation behavior studies, which this work is not within the scope of.

EquipmentList: This field is seldom completed, typically not formatted, and non-standard and hence unsuitable for regular analysis.

BookableUnitVirtual_ChildBookableUnitId: This field is utilized to represent virtual hierarchies between rooms or units, which are not directly used to measure utilisation.

The resulting working dataset thus contains a distilled set of features suitable to analysis across time (e.g., date starts and ends), space (e.g., rooms and buildings), and organisational patterns (e.g., bookings for teams and services). They are variables that enable full analysis of the use of space by dimension.

3.6 Assumptions

There are key assumptions that underpin the interpretation of space utilisation patterns within the NHS estate. These assumptions ensure consistency in data processing and enable meaningful insights into booking behaviour and room usage. They are as follows:

1) All bookings that were not cancelled were utilised.

2) All cancelled bookings remain cancelled

3) The average building opening and closing time are 8:00 and 18:00 respectively

3.7 Ethical considerations

Below are some ethical issues related to this study;

Staff Confidentiality and Privacy

Although the dataset is regarding bookings of resources and rooms rather than patient treatment, there will be bookings that have sensitive information relating to staff work patterns, shifts, or department activity. Ethical handling will require anonymisation and aggregation so that no individual staff can be identified from booking history.

Data Security and Governance

NHS data, when being used, must be processed securely and in compliance with GDPR and NHS data governance rules. These include secure storage, access control, and careful recording of all preprocessing steps for ensuring data integrity and traceability.

Transparency of Analysis

Spare building or bookable space decisions may affect staff flexibility, office convenience, or access to facilities. Transparency and reproducibility, to be equitable, must be the rules governing the evaluation process, with results openly reported to stakeholders.

Fairness and Equity for Staff

Downsizing or decommissioning resources according to utilisation has the potential to disadvantage the staff in particular districts or departments. Ethical consideration requires a consideration of whether less availability can lead to inequality of workload, utilisation of meeting space, or support for provision of service.

Responsible Use of Forecasting

Although predictive models are useful in providing some guidance, they should be used as decision making tools and not as guarantees. Over reliance on forecasts without consulting the stakeholders may lead to unnecessary reductions in resources that cause detriment to staff functioning and morale.

Preventing Structural Bias

Historical patterns of use may be employed as an indicator of structural issues like uneven resource distribution or access problems. Professional ethics require one to recognise that "low use" does not always equate with "low need" and consider context before any final decisions are made.

CHAPTER 4

IMPLEMENTATION, EVALUATION, RESULTS AND DISCUSSION

4.1 Implementation and evaluation

Below is the breakdown of the implementation process for the NHS space utilisation project alongside the corresponding code snippets and visualisations;

4.1.1 Data collection

As discussed in chapter three, the booking dataset for this study is a consolidated dataset of data extracts from the National Health Service (NHS) booking systems that capture different sections of the bookings from 2016 to 2026(future bookings inclusive) and provides a rich tapestry of data on the usage of desks and meeting rooms across diverse range of districts containing buildings that house the bookable units.

4.1.2 Data preprocessing

Merging

As discussed in chapter three, the merging process followed a systematic approach as summarised in table 2.1 and table 2.2 below. Table 2.1 shows the merging process for the individual datasets into seven (7) steps while table 2.2 shows the merging process for the seven (7) steps to form the consolidated booking dataset;

		TABLE JOINS AND RELATIONSHIP 1			
Step	**Table A**	**Join Column A**	**Table B**	**Join Column B**	**Purpose**
Step 1	Booking	Booking_TeamId	Team	Team_Id	Get team details for the booking
	Team	Team_ServiceId	Service	Service_Id	Get service associated with the team
	Service	Service_BusinessDeliveryUnitId	BusinessDeliveryUnit	BusinessDeliveryUnit_Id	Get business delivery unit info for the service
Step 2	PersonBooking	PersonBooking_PersonId	Person	Person_Id	Get person (user) details for the booking
	PersonBooking	PersonBooking_PersonBookingTypeId	PersonBookingType	PersonBookingType_Id	Get type of person involvement (e.g., Organizer, Participant)
Step 3	Booking	Booking_BookableUnitId	BookableUnit	BookableUnit_Id	Get physical space/unit booked
	BookableUnit	BookableUnit_AreaId	Area	Area_Id	Get area information for the bookable unit
	Area	Area_BuildingId	Building	Building_Id	Get building where the area is located
	Building	Building_DistrictId	District	District_Id	Get district where the building is located
Step 4	BookableUnit	BookableUnit_Id	BookableUnitEquipment	BookableUnitEquipment_BookableUnitId	Get list of equipment linked to the bookable unit
	BookableUnitEquipment	BookableUnitEquipment_EquipmentId	Equipment	Equipment_Id	Get actual equipment titles
Step 5	BookableUnit	BookableUnit_BookableUnitTypeId	BookableUnitType	BookableUnitType_Id	Get type of bookable unit (e.g., Meeting Room, Desk)
	BookableUnit	BookableUnit_Id	BookableUnitVirtual	BookableUnitVirtual_BookableUnitId	Get virtual details like child units and capacity
Step 6	BookingOccasion	BookingOccasion_CancellationReasonId	CancellationReason	CancellationReason_Id	Get reason if the occasion was canceled
	BookingOccasion	BookingOccasion_BookingOccasionStateId	BookingOccasionState	BookingOccasionState_Id	Get current state of the occasion (e.g., Confirmed, Cancelled)
Step 7	Booking	Booking_BookingTypeId	BookingType	BookingType_Id	Get type of booking (e.g., Internal, External)
	Booking	Booking_BookingStateId	BookingState	BookingState_Id	Get current booking status (e.g., Active, Inactive)
	Booking	Booking_BookingAuthorisationStatusId	BookingAuthorisationStatus	BookingAuthorisationStatus_Id	Get authorization status (e.g., Approved, Pending Approval)

Table 2.1 : Merging table1

		TABLE JOINS AND RELATIONSHIP 2			
Join Step	**From Table**	**Join Column**	**To Table**	**Join Column**	**Purpose**
1	Step7_Booking_Metadata	Booking_Id	Step1_Booking_Org	Booking_Id	Add team, service, and business delivery unit info
2	Step7_Booking_Metadata	Booking_Id	Step2_Booking_People	PersonBooking_BookingId	Add person and role details for each booking
3	Step7_Booking_Metadata	Booking_Id	Step3_Booking_Location	Booking_Id	Add location details (unit, area, building, district)
4	Step3_Booking_Location	BookableUnit_Id	Step4_Booking_Equipment	BookableUnit_Id	Add list of equipment available in the booked unit
5	Step3_Booking_Location	BookableUnit_Id	Step5_Booking_UnitTypes	BookableUnit_Id	Add unit type (e.g., Meeting Room), virtual child units and capacity
6	Step7_Booking_Metadata	Booking_Id	Step6_Booking_Occasion	BookingOccasion_BookingId	Add occasion details like start/end time, cancellation reason, state

Table 2.2 : Merging table2

SQL merging scripts

To merge, all the extracts datasets were first uploaded into the SQL environment before proceeding with the merging steps as seen in the merging flowchart in chapter three. See appendix A for screenshots of full SQL merging operation.

After consolidating the final merged dataset, it is then exported from SQL environment, saved as an excel file and renamed as BOOKING DATASET.

Following the dataset merge and export, the information were thoroughly cross checked with the original information contained in the individual datasets especially the booking occasion dataset to verify that the merging process yielded the right result without mismatching information to avoid altering the credibility of the research results and recommendations. After verifying the reliability of the merged dataset, it is then uploaded on google colab for further analysis as seen in fig. 5.1 and fig. 5.2. See full information in appendix B.

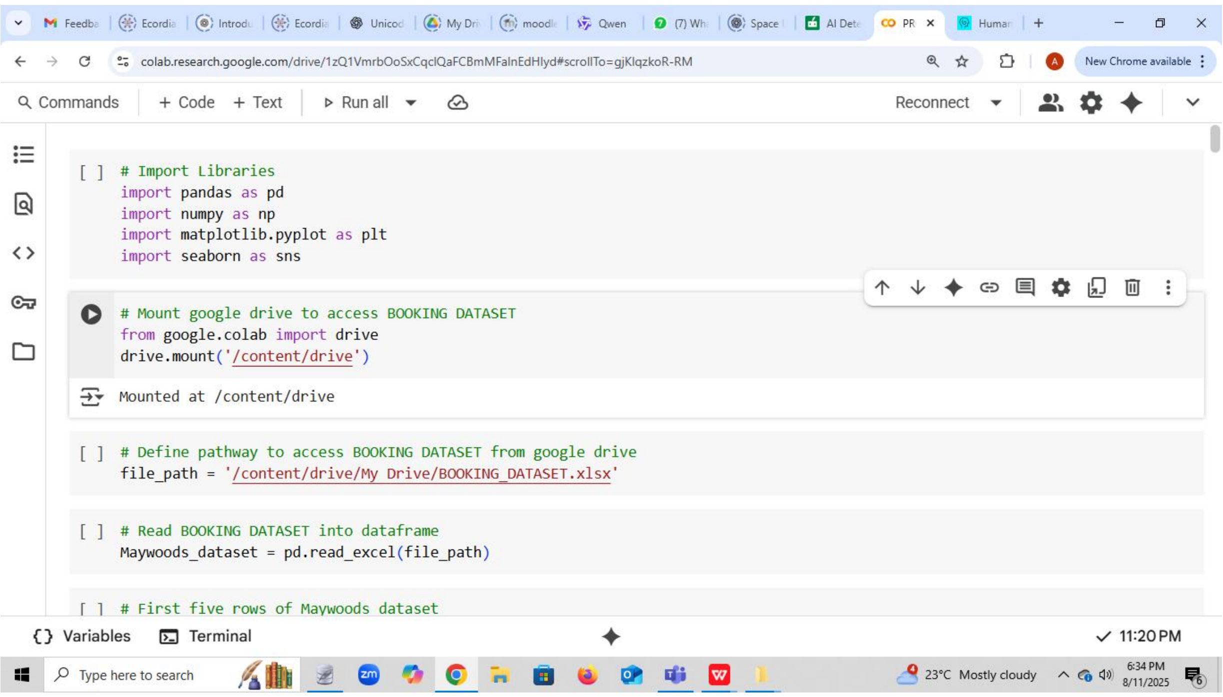


Fig. 5.1: Upload of Booking dataset on google colab

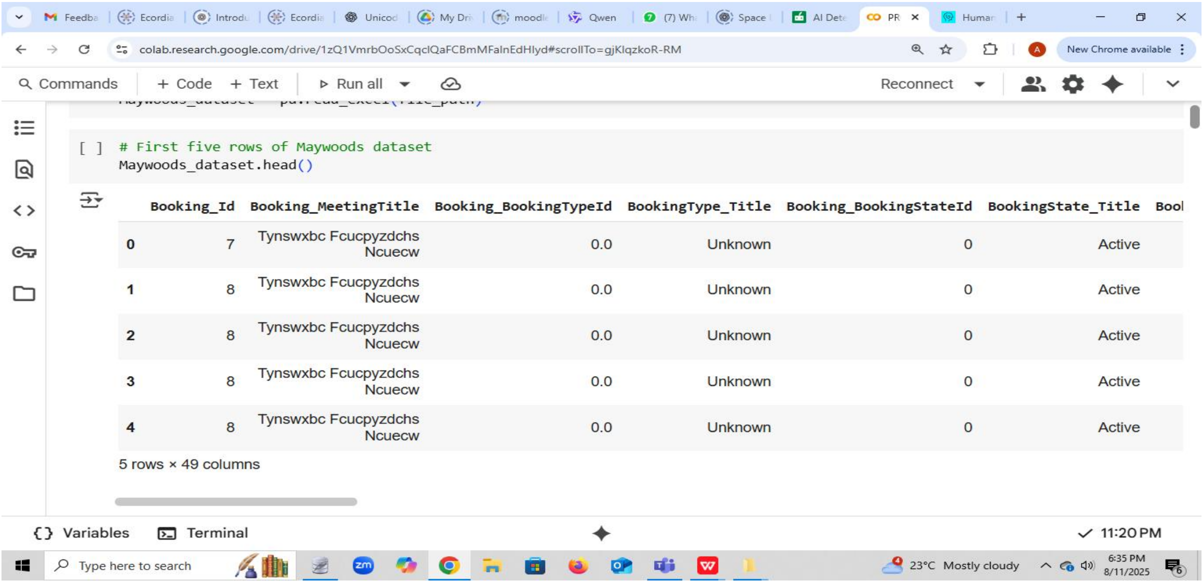


Fig. 5.2: Showing first 5 rows of Maywoods dataset

Data summary

The merged maywoods dataset contains 43 variables and 1048576 entries which include relevant dimensions such as timing information (booking and building start and end times), space identifiers (bookable unit titles, building IDs, and district names etc), user information (team and service information), and metadata about bookings (booking status, authorisation levels, number of visitors etc).as shown in the fig. 6.1 below. See full information in appendix B.

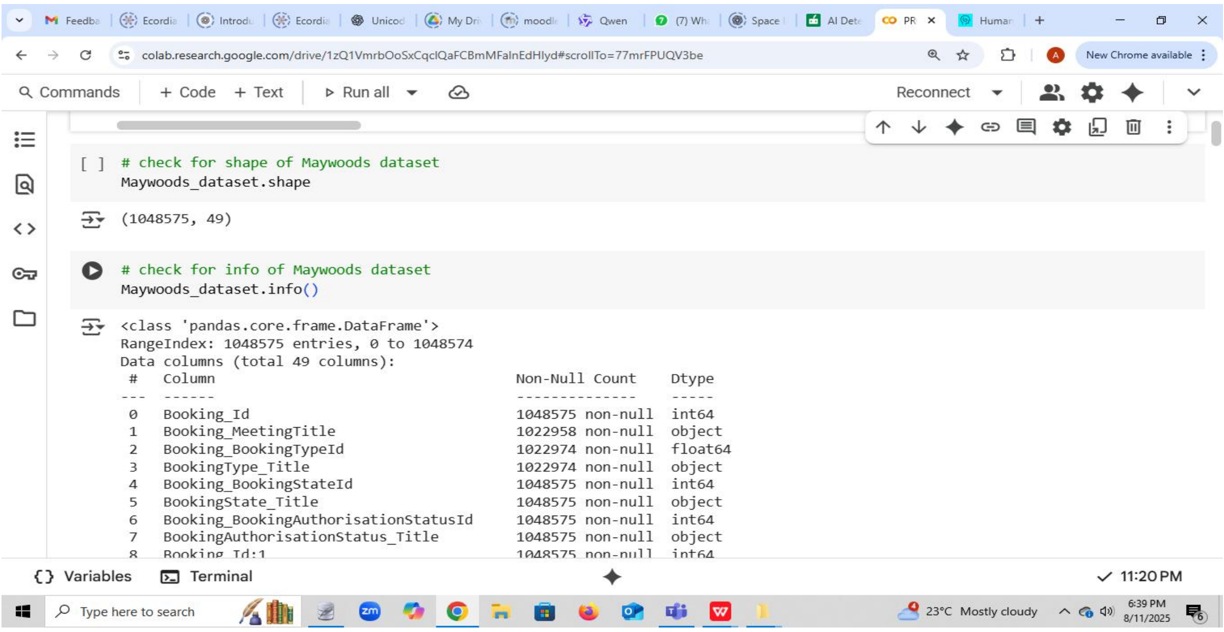


Fig. 6.1: Maywoods dataset shape and info

Drop duplicated columns

The merging of the various booking datasets that formed the maywoods dataset resulted to some columns being duplicated as listed below;

Booking_Id:1,

Booking_Id:2,

PersonBooking_BookingId,

BookingOccasion_BookingId,

BookableUnit_Id:1,

BookableUnit_Id:2,

As part of the data cleaning process, the aforementioned columns are deduplicated as seen in figure …. Note that duplicated rows have been dropped in the merging stage in SQL.

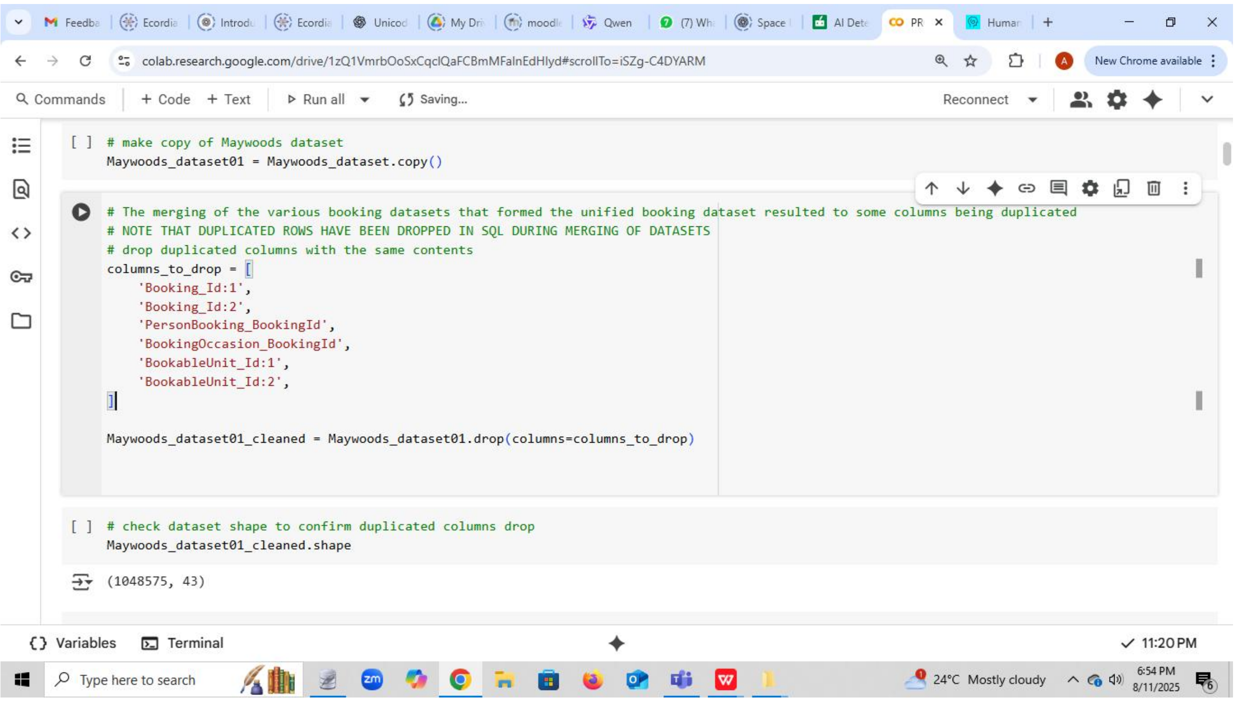


Fig. 7: Drop duplicated columns

Check for null values

A summary of the check for null values in the maywoods dataset is seen in fig.8.1 below. See appendix B for full information.

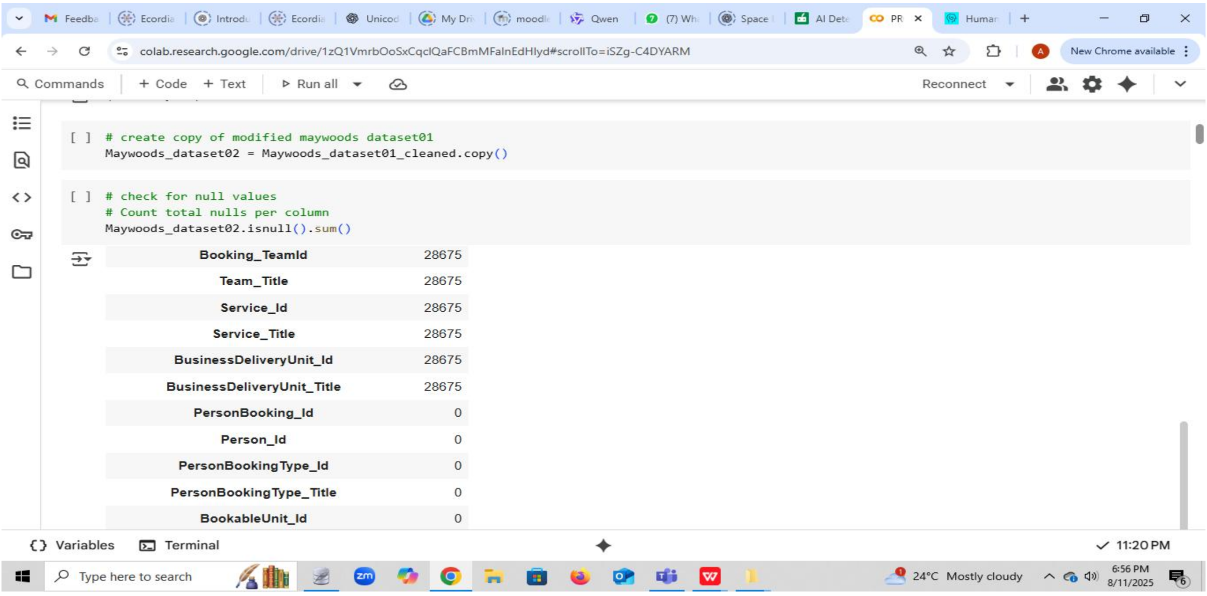


Fig.8.1: Check for null values_1

Convert date columns to datetime

The date columns of the datasets are converted to datetime for better analysis approach as shown in figure .....

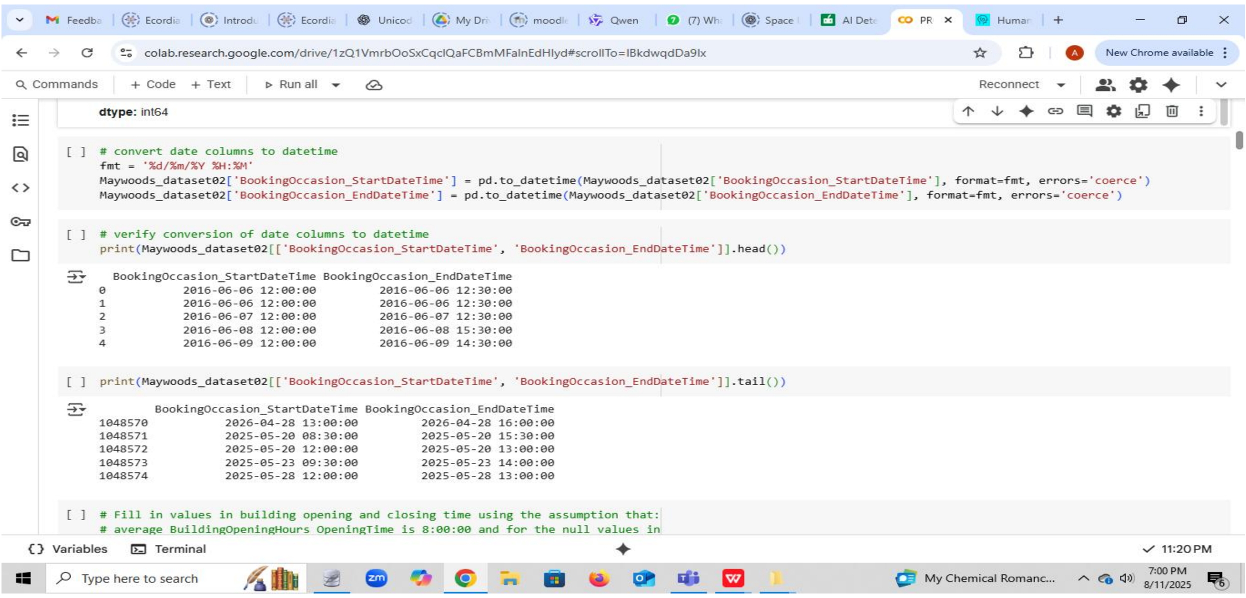


Fig.9: Convert date column to datetime

Handling null values

For better insight into the temporal pattern of the dataset, the null values for the building opening and closing time are handled as shown in fig.10 below using assumption three as stated in chapter three that the average opening and closing time of a building is 8:00 and 18:00 respectively

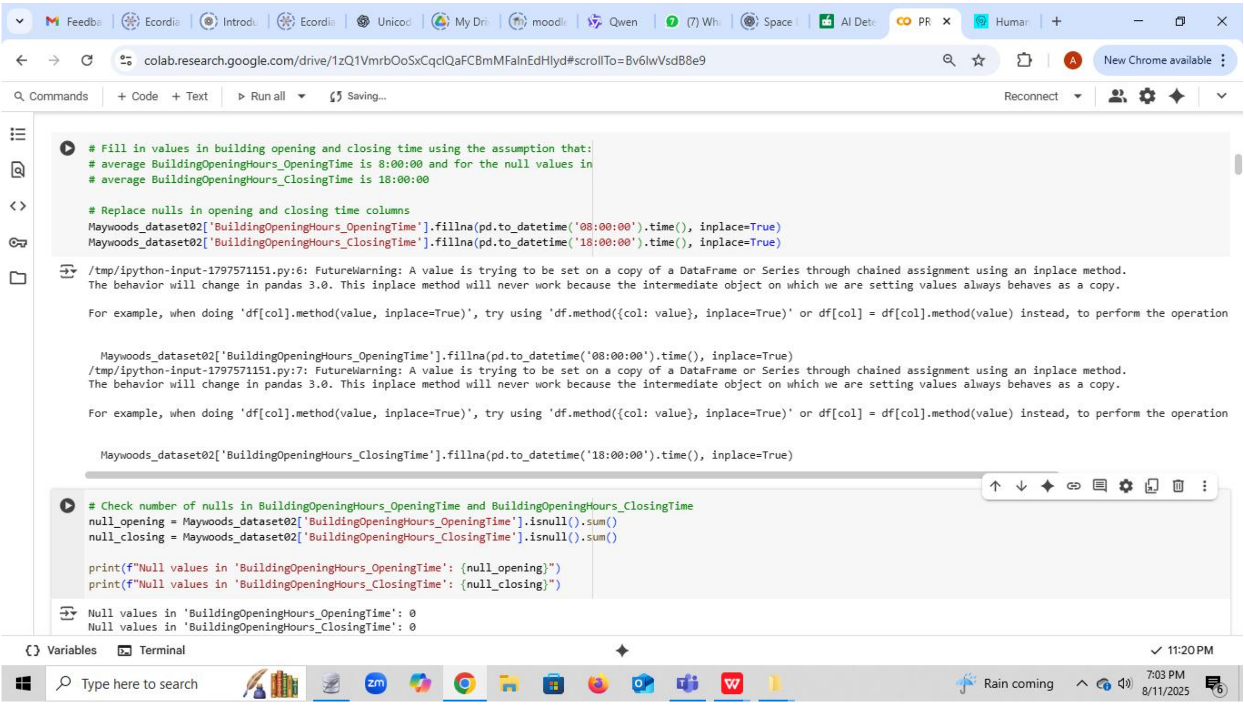


Fig. 10: Handle null values of Building opening and closing hours using assumption 3

Check for outliers

In alignment with the scope of this study, the outliers check using boxplot is focused on the building opening and closing hours in order to ensure they fall within the expected duration to eradicate any form of bias in the utilisation analysis. As seen in fig. 11b and fig. 11c below, there are no observable outliers in the building opening and closing hours as most of the buildings open around 8:00 with very few opening around 7:00 and 9:00 as expected. And for the building closing hours, there are no outliers as most of the buildings close around 18:00 with few closing around 17:00, 19:00 and 20:00 as expected. See appendix B (fig. 11a) for full code.

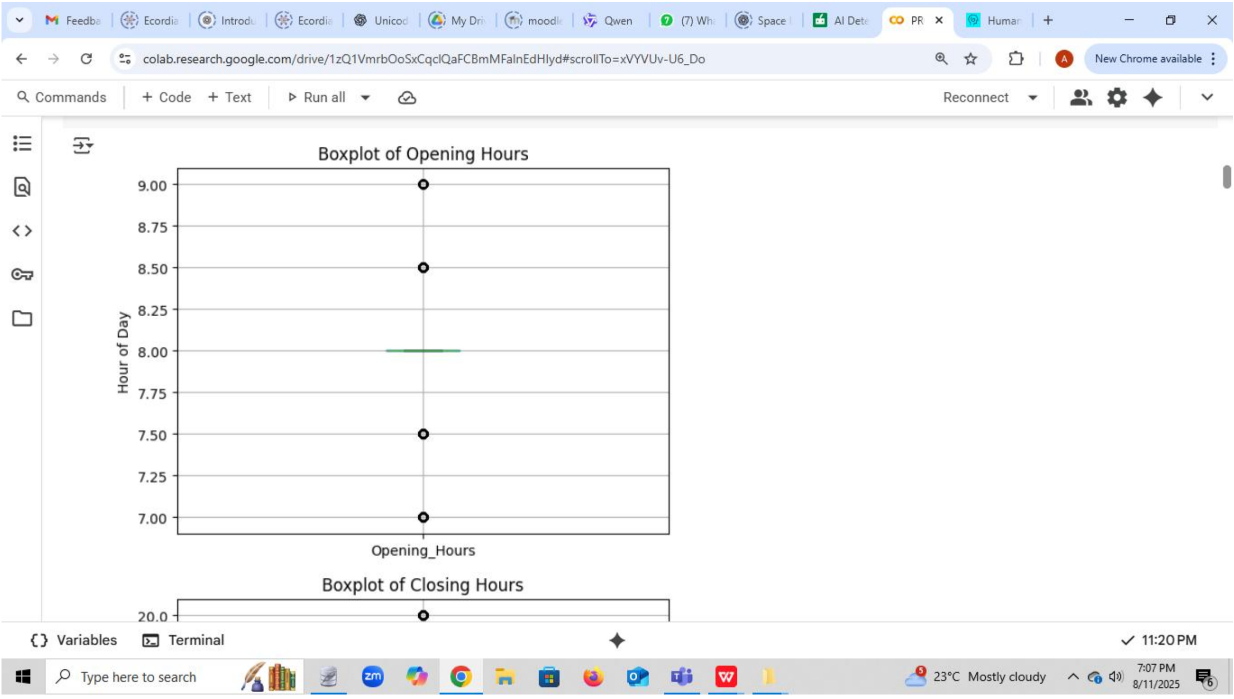


Fig. 11b: Check for outliers (Building opening hours)

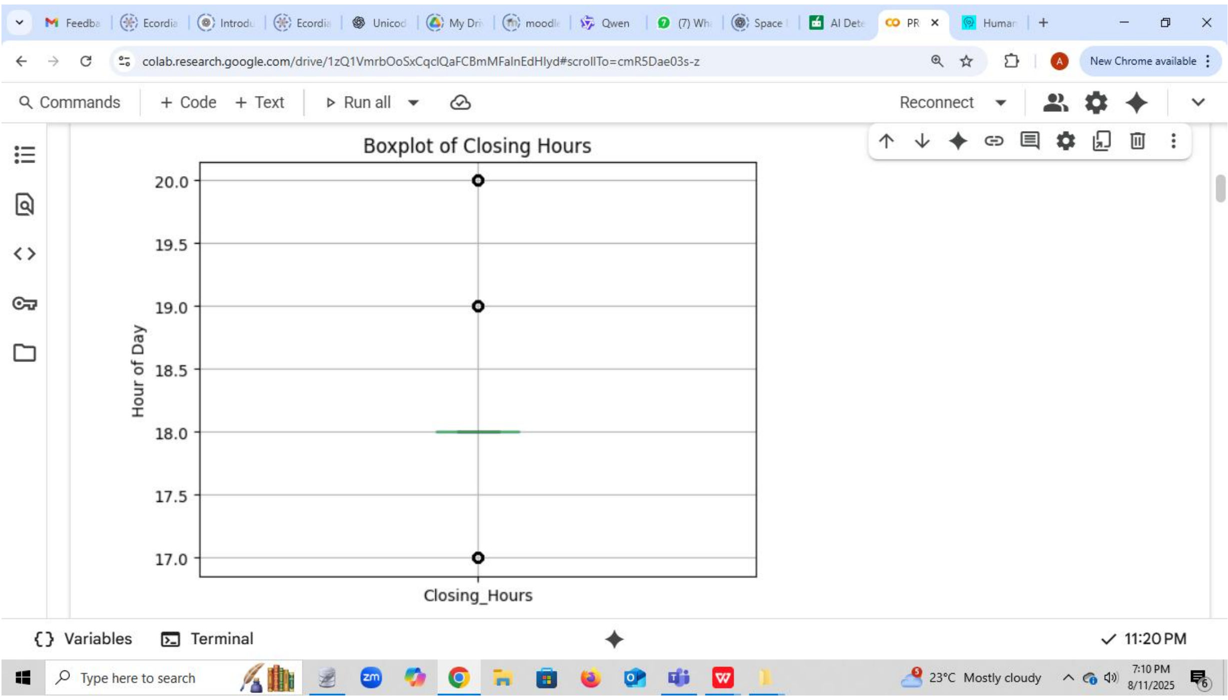


Fig. 11c: Check for outliers (Building closing hours)

Tabular and visual summary of Maywoods dataset before evaluation

Below is the visual summary of the maywoods dataset in bar charts and tables with respect to districts, buildings and bookable units before evaluation. The dataset is made up of booking information from five (5) districts containing twenty three (23) buildings with a total of four hundred and three (403) bookable units. Table 3 shows the district ids alongside the corresponding building ids and the bookable unit count for each building id while fig. 12.2(b) and fig. 12.3(b) show the visual summary. See full code in appendix B.

District_id	Building_id	Bookable unit count
3	4	4
4	5	40
	6	21
	19	84
	22	3
	24	11
	39	4
	46	6
	47	1
5	7	25
	8	19
	17	1

	20	1
	25	17
6	9	12
	10	45
	11	20
	12	55
	13	10
	14	3
	35	1
	36	3
7	48	17
Total count **5**	**23**	**403**

Table 3: Booking dataset summary before evaluation

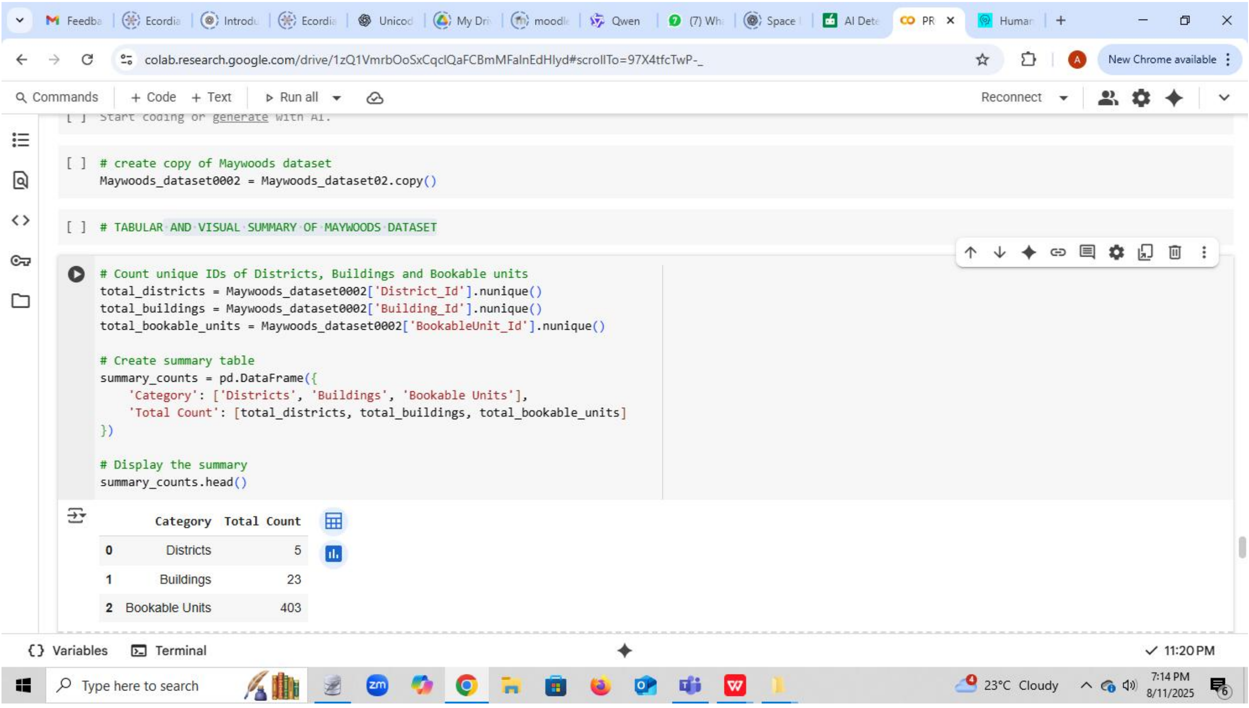


Fig. 12.1: Tabular summary of Maywoods dataset

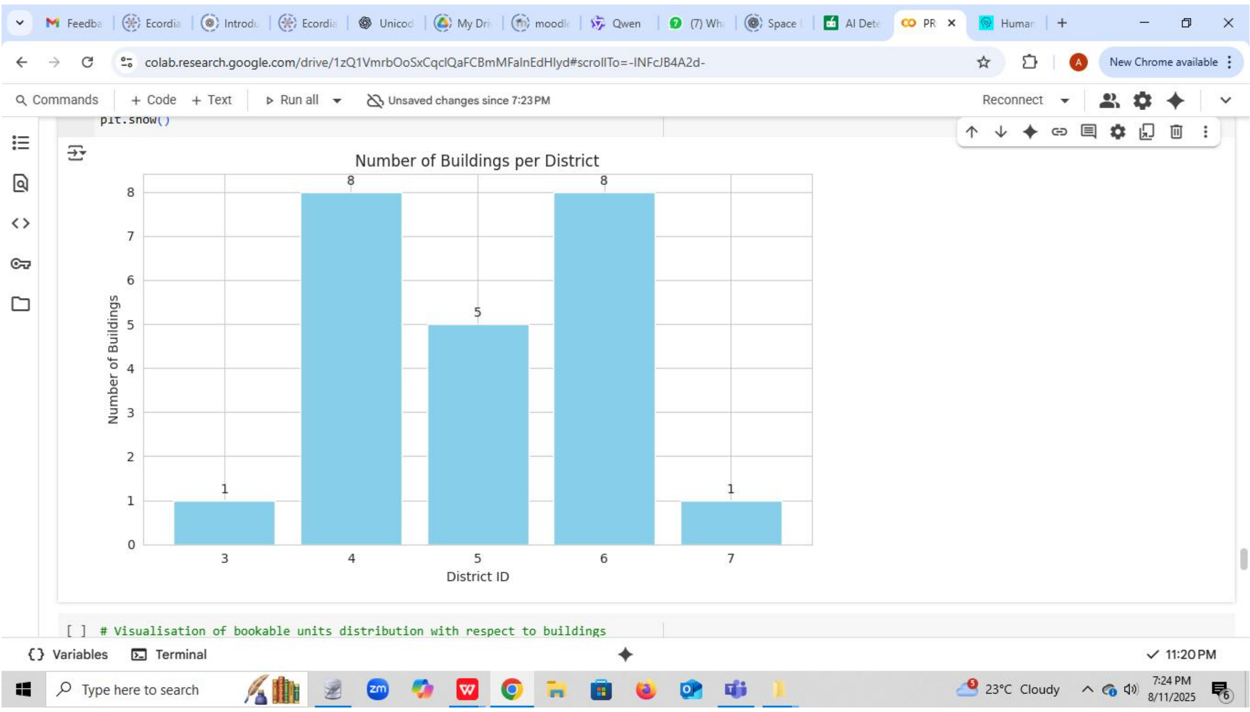


Fig. 12.2(b): Visualisation of buildings distribution with respect to districts 2

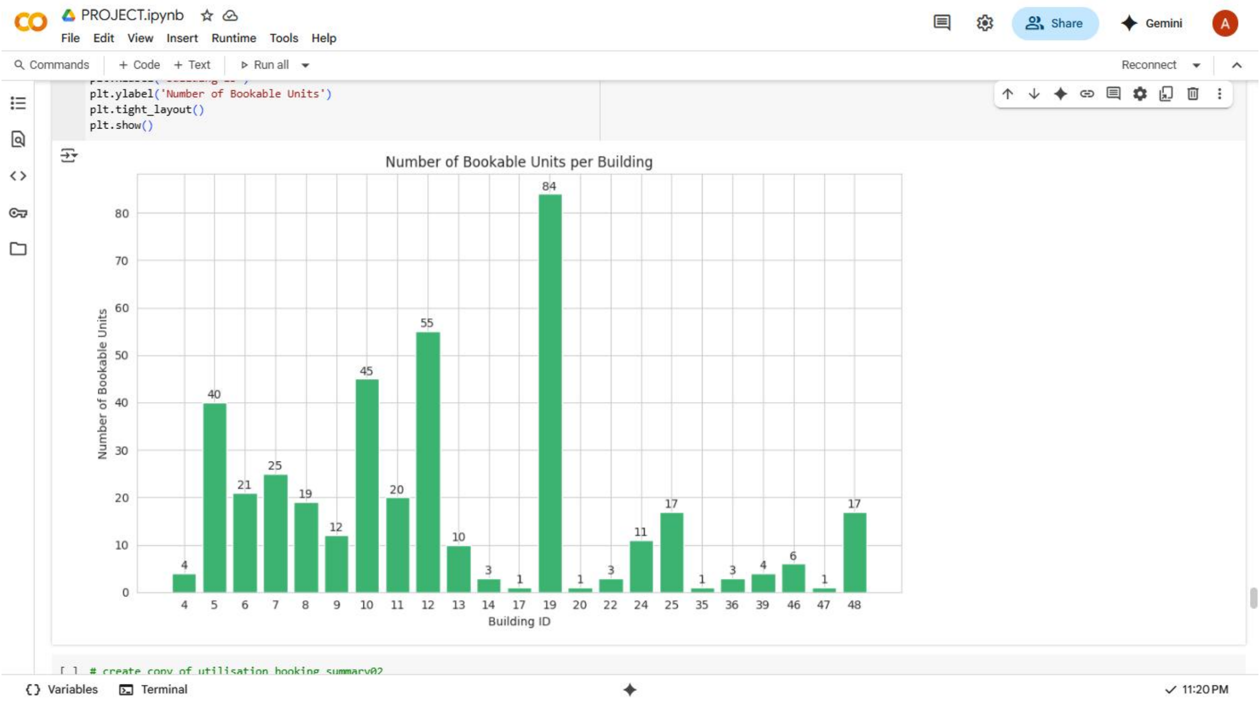


Fig. 12.3(b): Visualisation of bookable units distribution with respect to buildings 2

Feature engineering

As mentioned in chapter three, some feature engineering were carried out. In the course of plotting the boxplot to check for outliers in the building opening and closing hours, two additional columns (Opening Hours and Closing Hours) were created. These columns are used to calculate and create the actual building operating time column by subtracting the opening hours from the closing hours as shown in fig. 13.2. To calculate the actual booking hours, the booking occasion start time is subtracted from the booking occasion end time to estimate the duration of each booking as shown in fig. 13.1.

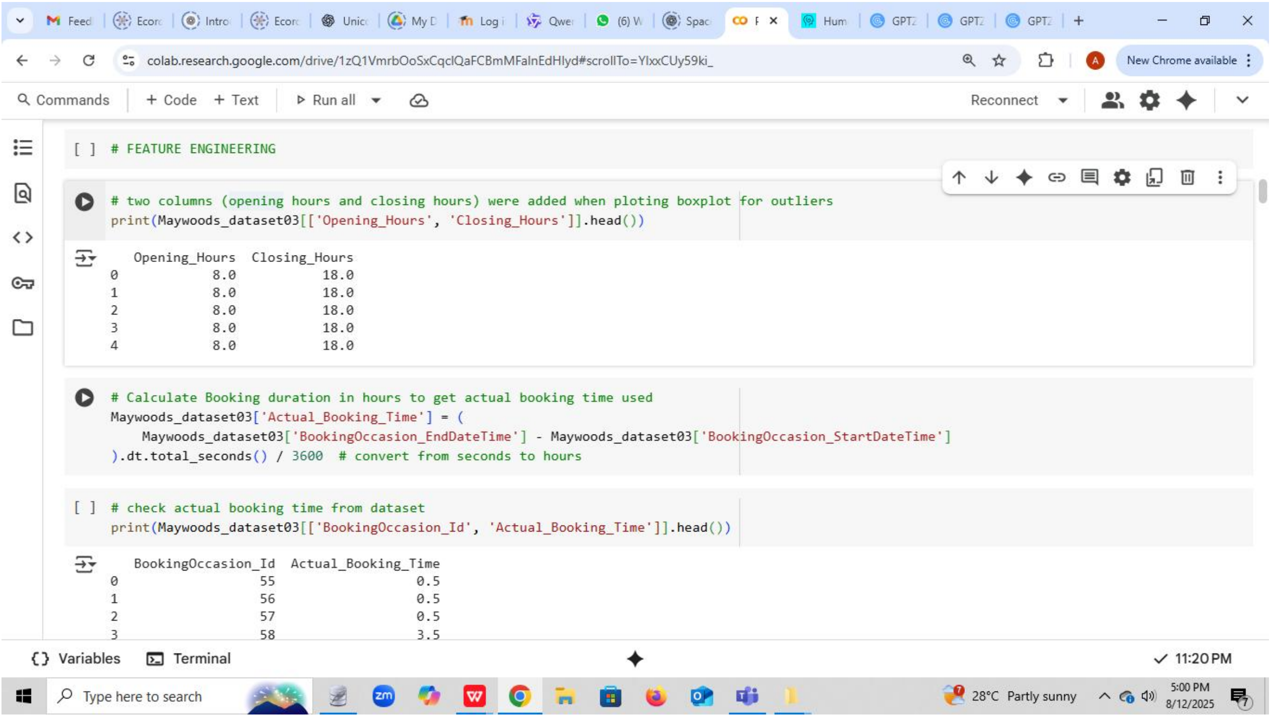


Fig. 13.1: Actual booking hours

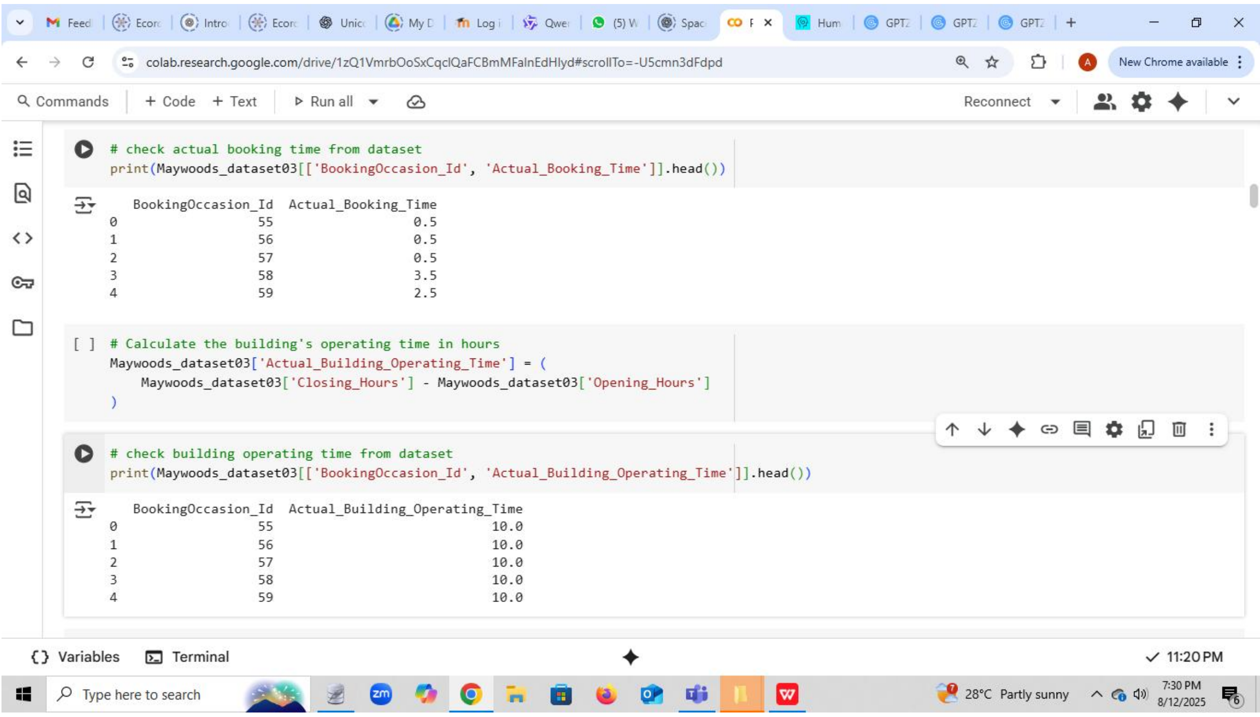
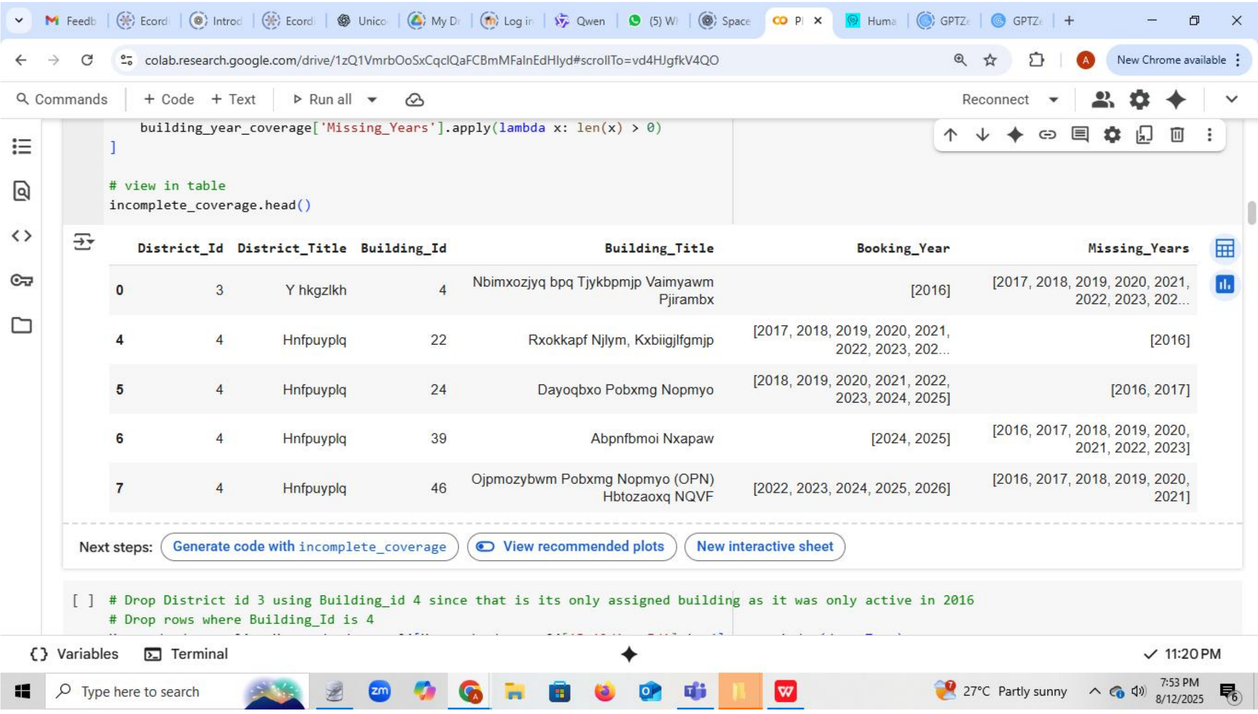


Fig. 13.2: Actual building operating hours

Data cleaning

Before analysis, the maywoods dataset is reviewed to check for district(s) with inconsistent booking activities i.e.district(s) not having up-to-date booking activities in order to drop them from the dataset and also bookings that were canceled using the cancellation reason column to check and drop the corresponding rows before evaluation as shown in fig. 13.2. See full code in appendix B. District 3 only recorded active bookings in 2016 as seen in the output thereby making it unfit to be included in the analysis. It is then dropped alongside its corresponding building as shown in appendix B likewise the canceled bookings.



```
    building_year_coverage['Missing_Years'].apply(lambda x: len(x) > 0)
]

# view in table
incomplete_coverage.head()
```

	District_Id	District_Title	Building_Id	Building_Title	Booking_Year	Missing_Years
0	3	Y hkgzlkh	4	Nbimxozjyq bpq Tjykbpmjp Vaimyawm Pjirambx	[2016]	[2017, 2018, 2019, 2020, 2021, 2022, 2023, 202...
4	4	Hnfpuyplq	22	Rxokkapf Njlym, Kxbiigjlfgmjp	[2017, 2018, 2019, 2020, 2021, 2022, 2023, 202...	[2016]
5	4	Hnfpuyplq	24	Dayoqbxo Pobxmg Nopmyo	[2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025]	[2016, 2017]
6	4	Hnfpuyplq	39	Abpnfbmoi Nxapaw	[2024, 2025]	[2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023]
7	4	Hnfpuyplq	46	Ojpmozybwm Pobxmg Nopmyo (OPN) Hbtozaoxq NQVF	[2022, 2023, 2024, 2025, 2026]	[2016, 2017, 2018, 2019, 2020, 2021]

Next steps: Generate code with incomplete_coverage | View recommended plots | New interactive sheet

```
# Drop District id 3 using Building_id 4 since that is its only assigned building as it was only active in 2016
# Drop rows where Building_Id is 4
```

Fig. 14.2: Check for inconsistent districts_2

Before cleaning, the maywoods dataset had 1,048,575 entries but after cleaning, the dataset entries reduced to 859,096 with number of bookable units dropping from 403 to 389.

When checking for inactive districts, a new column (Booking Year) was added to the dataset as shown in fig. 15 below;

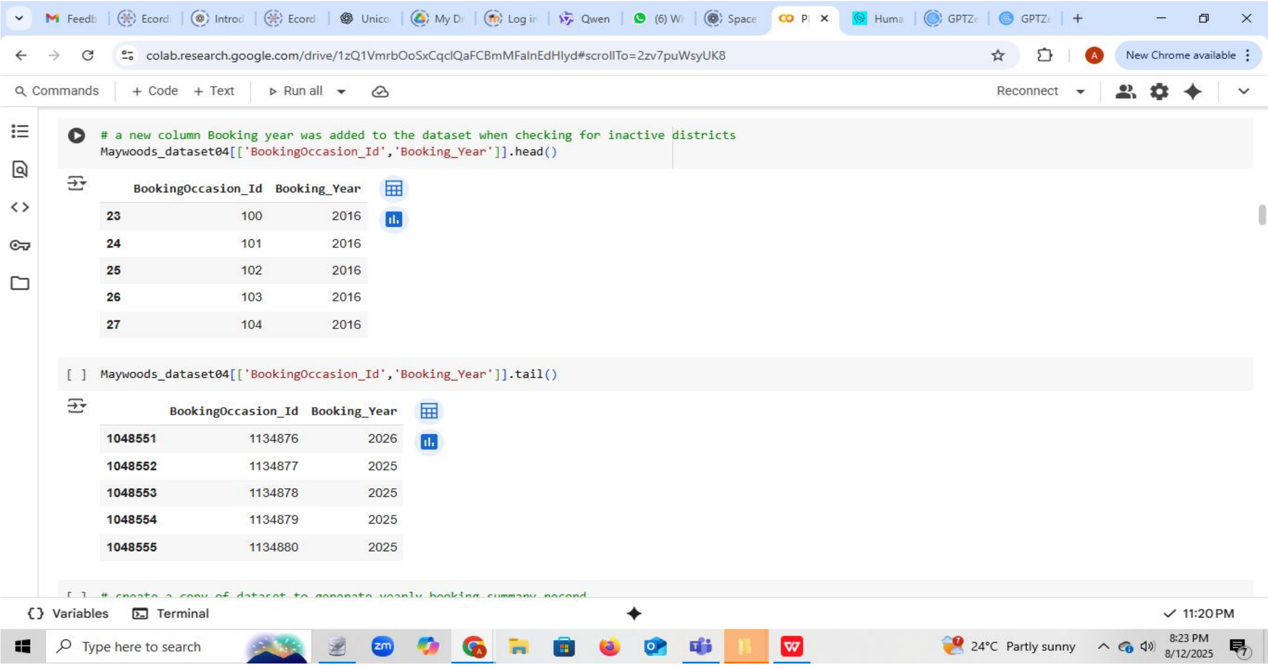


Fig. 15: Addition of Booking year column

4.1.3 Exploratory data analysis (EDA)

Yearly booking summary

To understand the booking history and pattern of the maywoods dataset, a yearly booking summary table is created that captures the yearly booking count of all bookable units from 2016 to 2026 as shown in fig. 16b and fig. 16c. See full code in appendix C (fig.16a).

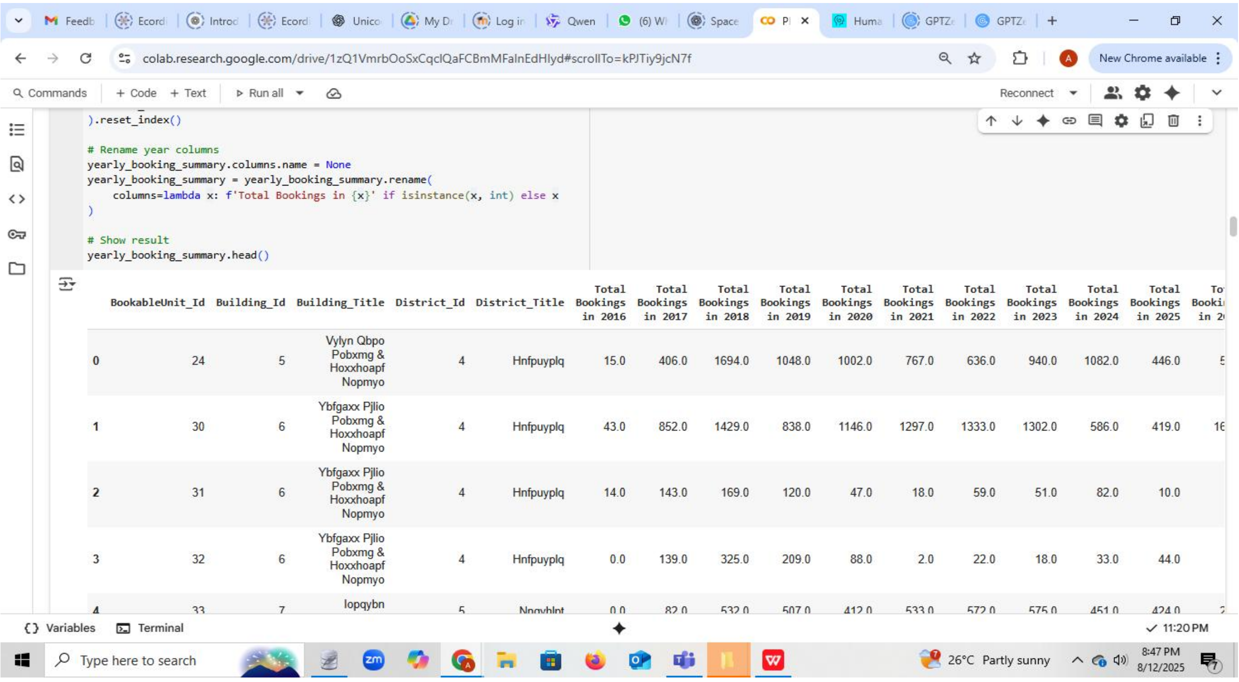


Fig. 16b: Yearly booking summary_2

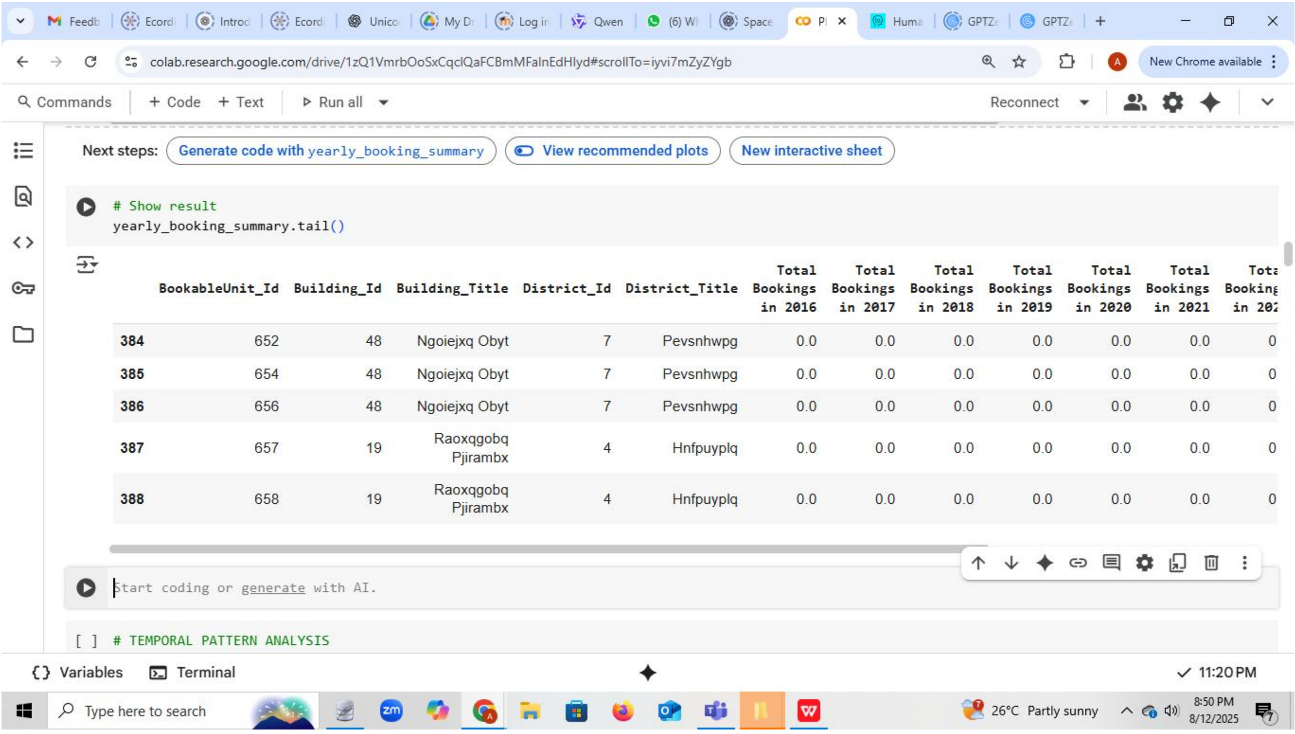


Fig. 16c: Yearly booking summary_3

4.1.3.1 Temporal pattern analysis

In this study, under temporal evaluation, line plot is used to examine how data changes over time to identify trends, patterns and behaviours with respect to yearly bookings, weekly bookings with focus on the days of the week with the highest and lowest bookings, and daily booking with focus on the hours of the day with the highest and lowest booking and actual booking hours. Below are the yearly, weekly, daily and actual duration booking temporal visual evaluation as shown in fig. 17.2 and 17.4, fig. 18.2, fig. 19.2 and fig. 20.2 respectively.

Fig. 17.2: Yearly booking line plot_2

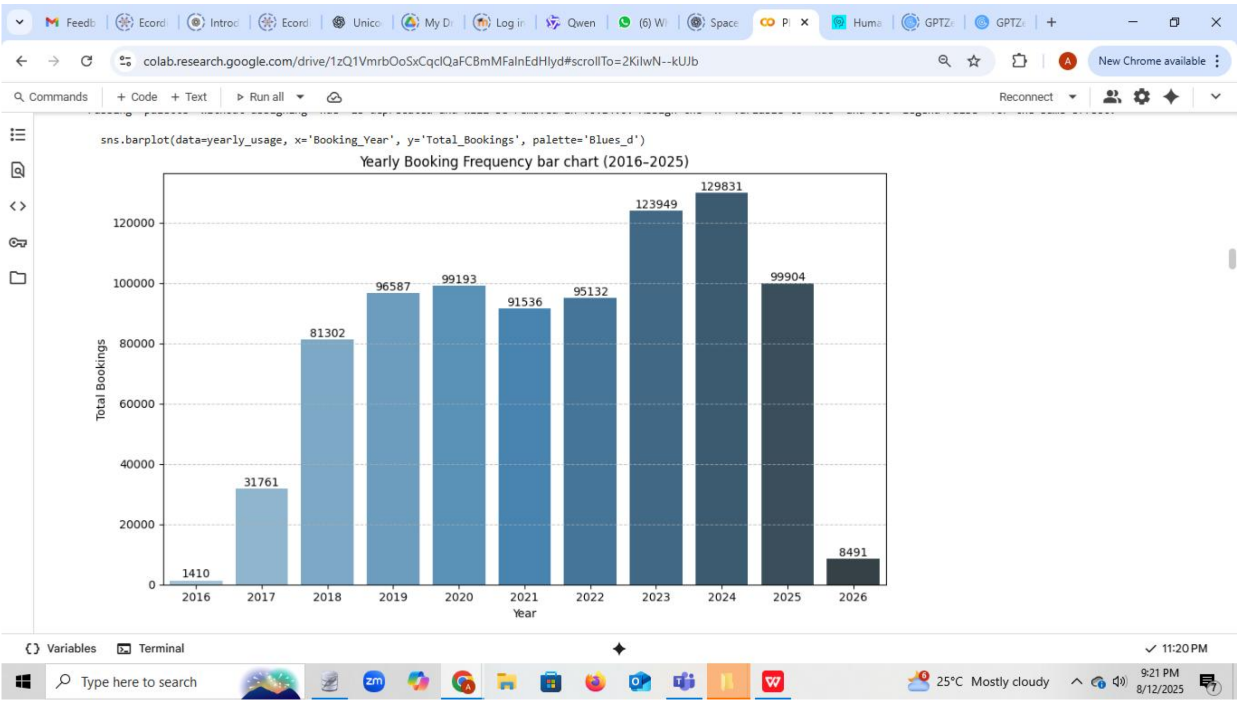


Fig. 17.4: Yearly booking bar chart_2

The year with the highest booking is 2024 with the lowest being 2016. Although the data captured till May of 2025 and from the trajectory 2025 shows a better chance of having the highest booking if the data is captured for the full year. See appendix C for full code (fig. 17.1 and 17.3).

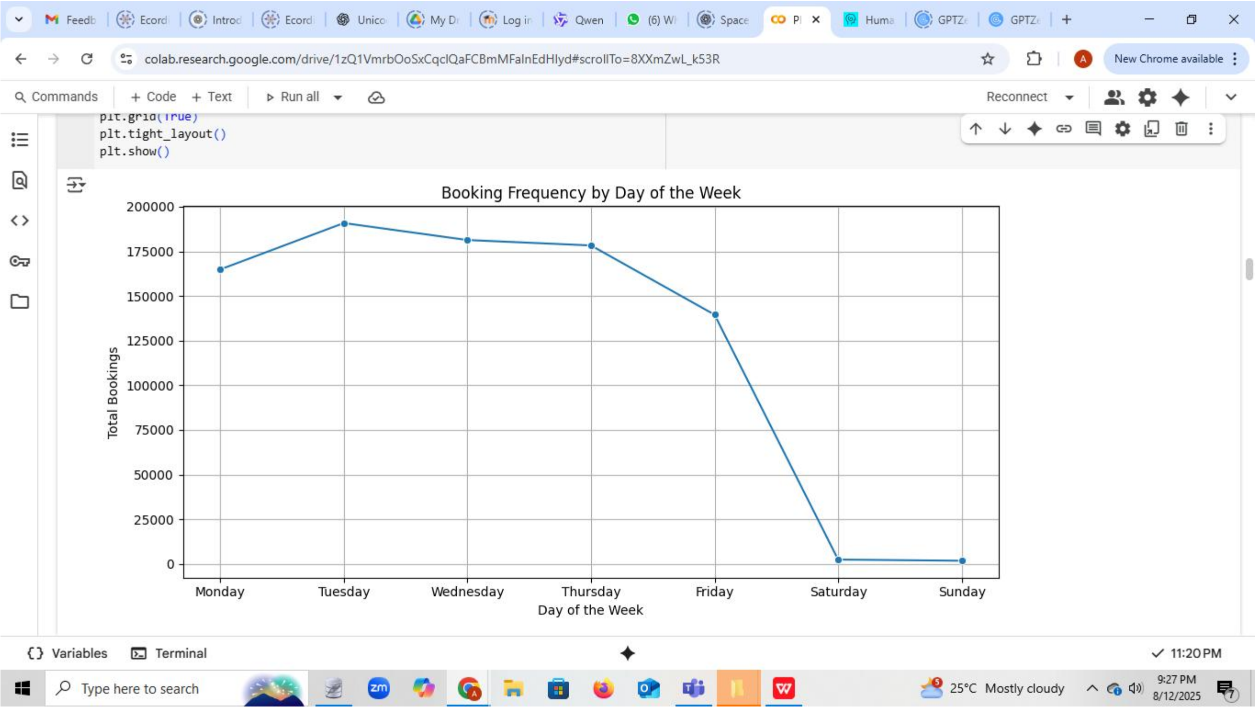


Fig. 18.2: Booking Frequency By Day Of The Week_2

From the booking frequency by day of the week, Tuesday have the highest with Saturday and Sunday having the lowest. See appendix C for full code (fig.18.1).

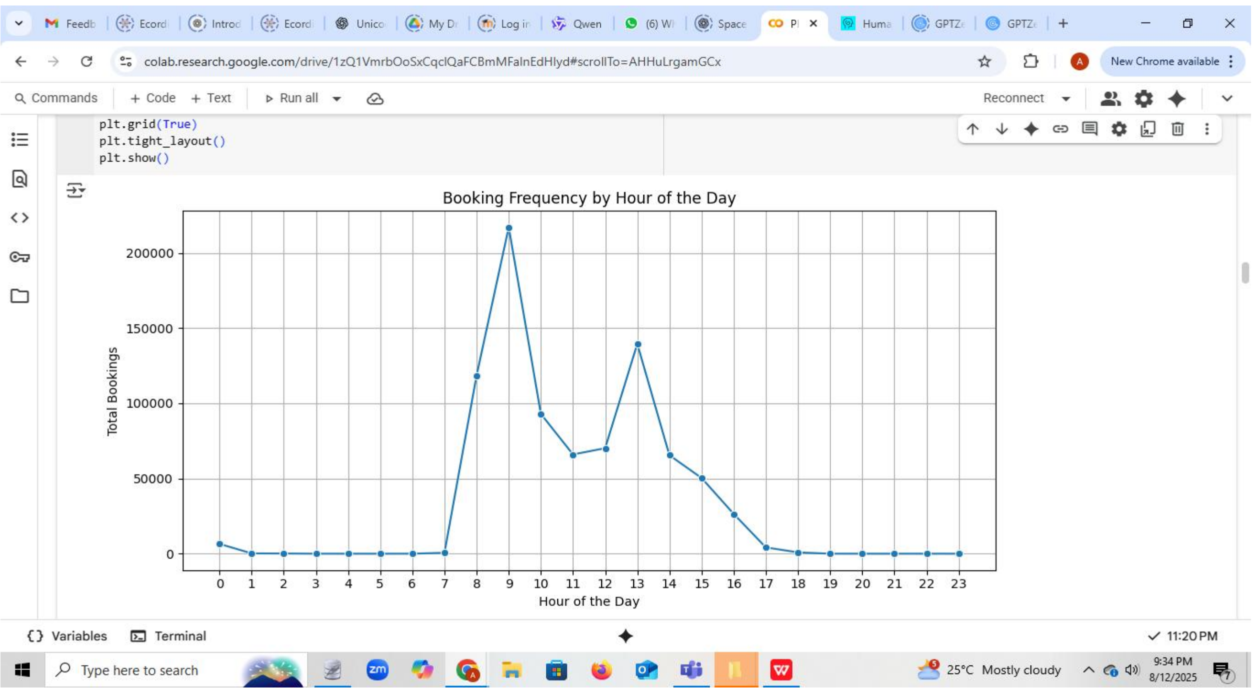


Fig. 19.2: Frequency distribution of Bookings by Hour of the Day_2

From the booking frequency by hour of the day, it can be seen that most bookings are done around 9:00 with the second highest being 13:00. See appendix C for full code (fig. 19.1).

Fig. 20.2: Booking Frequency By Actual Booking time used_2

From the actual booking plot, the duration for most meetings is one hour which can be explained for short team and service meetings with the second highest being eight hours which captured most likely training period. See appendix C for full code (fig. 20.1).

Temporal analysis summay

From the temporal patterns visual analysis, it can be seen that;

I) The year with the highest booking is 2024 with the lowest being 2016. Although the data captured till May of 2025 and from the trajectory 2025 shows a better chance of having the highest booking if the data is captured for the full year.

II) From the booking frequency by day of the week, Tuesday have the highest with Saturday and Sunday having the lowest.

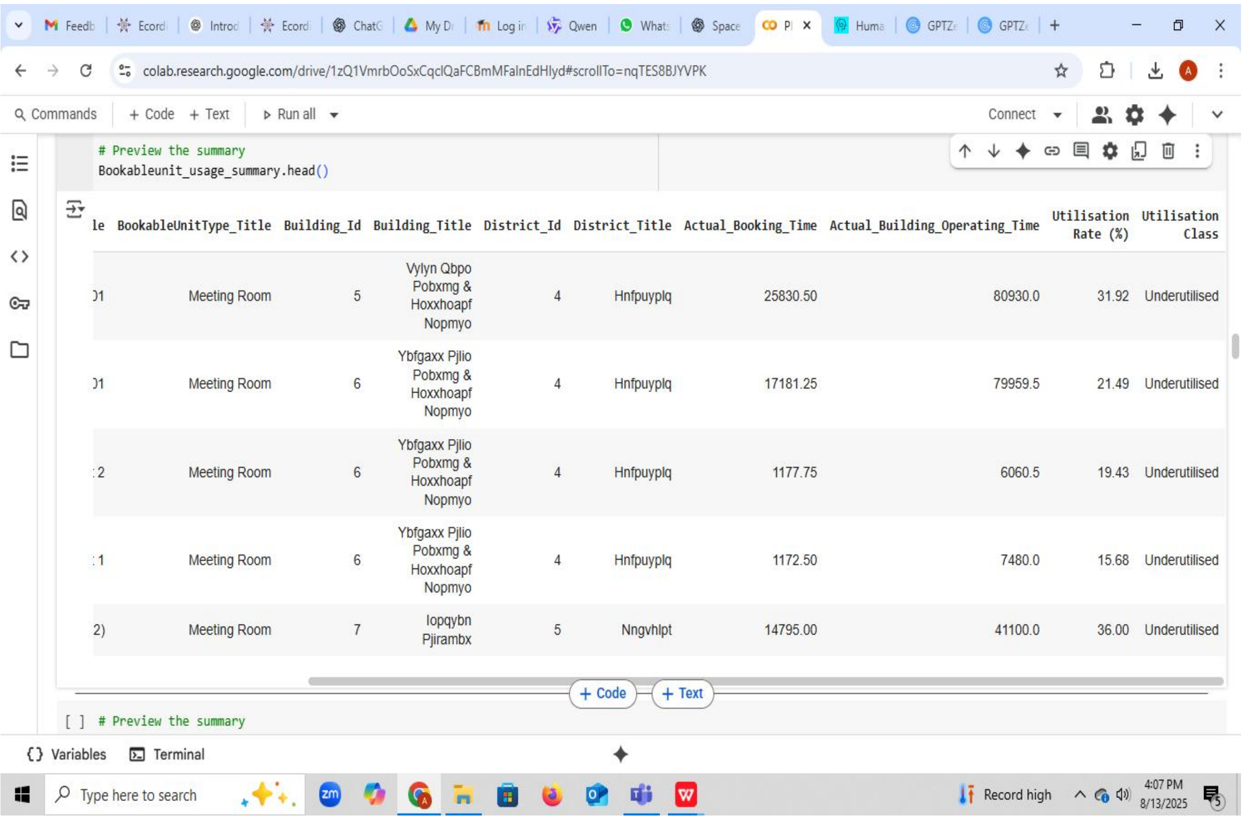
III) From the booking frequency by hour of the day, it can be seen that most bookings are done around 9:00 with the second highest being 13:00.

iv) From the actual booking plot, the duration for most meetings is one hour which can be explained for short team and service meetings with the second highest being eight hours which captured most likely training period.

4.1.3.2 Usage Analysis (Utilisation rate)

The utilisation rate is calculated in three categories (I) Bookable unit utilisation rate (ii) District utilisation rate and (iii) Building utilisation rate.

The bookable unit utilisation summary as seen in fig. 21.2(a) and 21.2(b) below shows the utilisation rate for all the bookable unit with a utilisation class categorising them into utilised for 50% utilisation rate and above and underutilised for utilisation rate below 50%. See full code in appendix C (fig. 21.1). In this study, the utilisation threshold to qualify a bookable unit to be categorised under utilised in the usage evaluation is 50% and above. This utilisation rate calculation is limited to the usage of the bookable unit during the time they were booked. The utilisation rate is gotten by dividing the actual booking time by the actual building operating time multiply by 100%.



le	BookableUnitType_Title	Building_Id	Building_Title	District_Id	District_Title	Actual_Booking_Time	Actual_Building_Operating_Time	Utilisation Rate (%)	Utilisation Class
)1	Meeting Room	5	Vylyn Qbpo Pobxmg & Hoxxhoapf Nopmyo	4	Hnfpuyplq	25830.50	80930.0	31.92	Underutilised
)1	Meeting Room	6	Ybfgaxx Pjlio Pobxmg & Hoxxhoapf Nopmyo	4	Hnfpuyplq	17181.25	79959.5	21.49	Underutilised
:2	Meeting Room	6	Ybfgaxx Pjlio Pobxmg & Hoxxhoapf Nopmyo	4	Hnfpuyplq	1177.75	6060.5	19.43	Underutilised
:1	Meeting Room	6	Ybfgaxx Pjlio Pobxmg & Hoxxhoapf Nopmyo	4	Hnfpuyplq	1172.50	7480.0	15.68	Underutilised
2)	Meeting Room	7	Iopqybn Pjirambx	5	Nngvhlpt	14795.00	41100.0	36.00	Underutilised

Fig. 21.2(a): Bookable unit utilisation summary_2

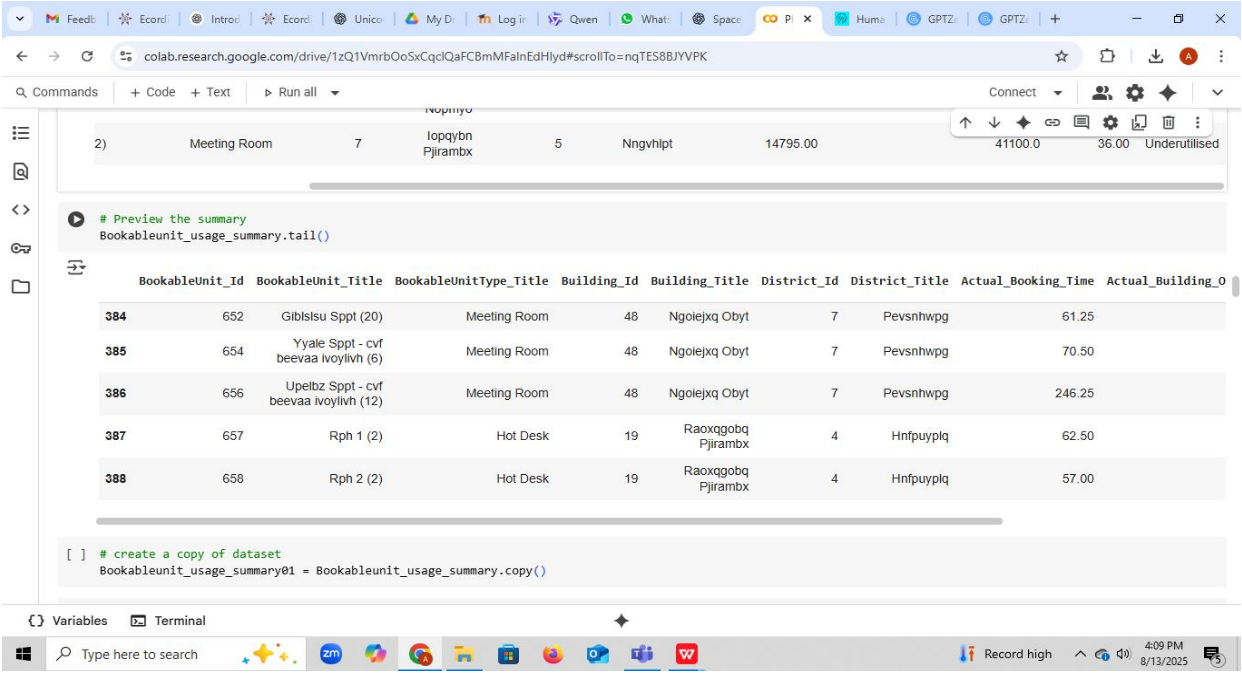


Fig. 21.2(b): Bookable unit utilisation summary_3

Frequency distribution for utilisation class

The utilisation class distribution as seen in fig. 22.2 shows that seventy nine (79) bookable units accounting for 20% of the total bookable units fall under the utilised class while 310 accounting for 80% fall under the underutilised class. The total bookable units dropped from 403 being the total in the merged data to 389 after data cleaning. See full code in appendix C (fig. 22.1).

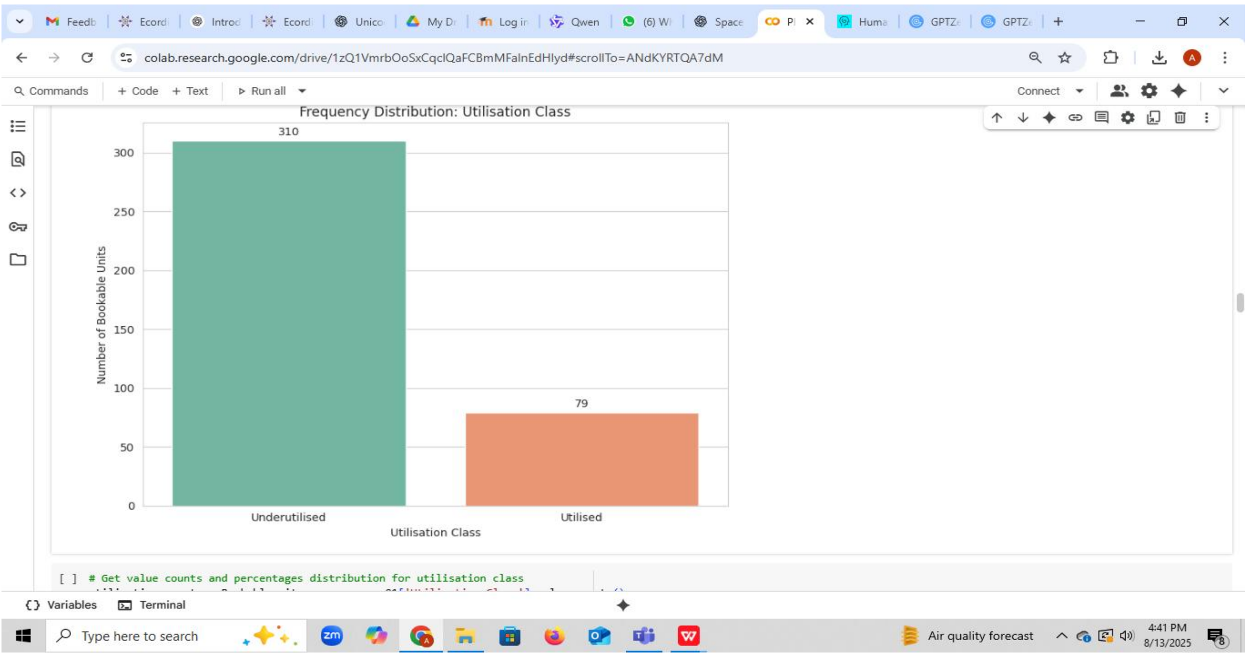


Fig. 22.2: Distribution for utilisation class_2

Clinical and non-clinical distribution of bookable units

The bookable units fall into two usage categories. It is either they are booked for clinical purpose or non-clinical purpose. Fig. 23.2. shows the frequency distribution of the bookable units with respect to usage type. It can be seen that 185 bookable units are for clinical purpose while 204 are for non-clinical purpose. See full code in appendix C (fig. 23.1).

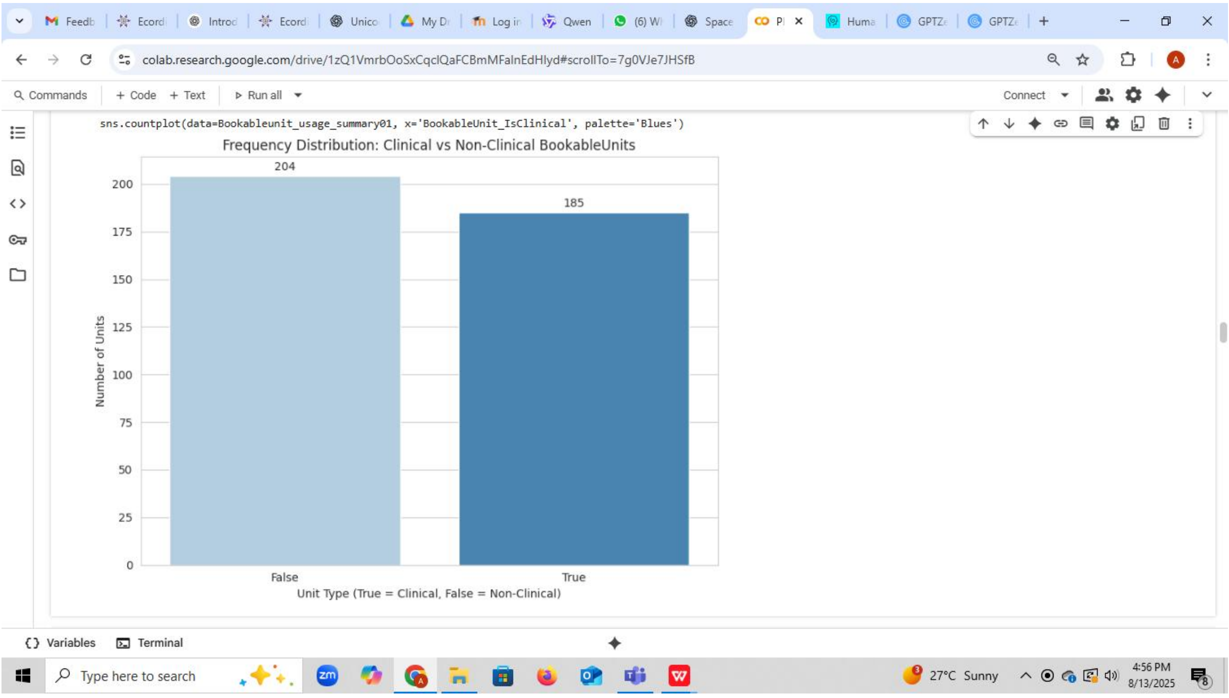


Fig. 23.2: Frequency distribution for clinical and non clinical 2

Districts utilisation rate

The district utilisation rate shows the average usage rate of the districts which is a function of the average cumulative usage of the buildings in each district as seen in fig. 24.2 and fig. 24.4. From the visual analysis, district id 4 has the highest utilisation rate of 36.22% while district id 7 has the lowest utilisation rate of 19.19%. See full code in appendix C (fig. 24.1 and fig. 24.3).

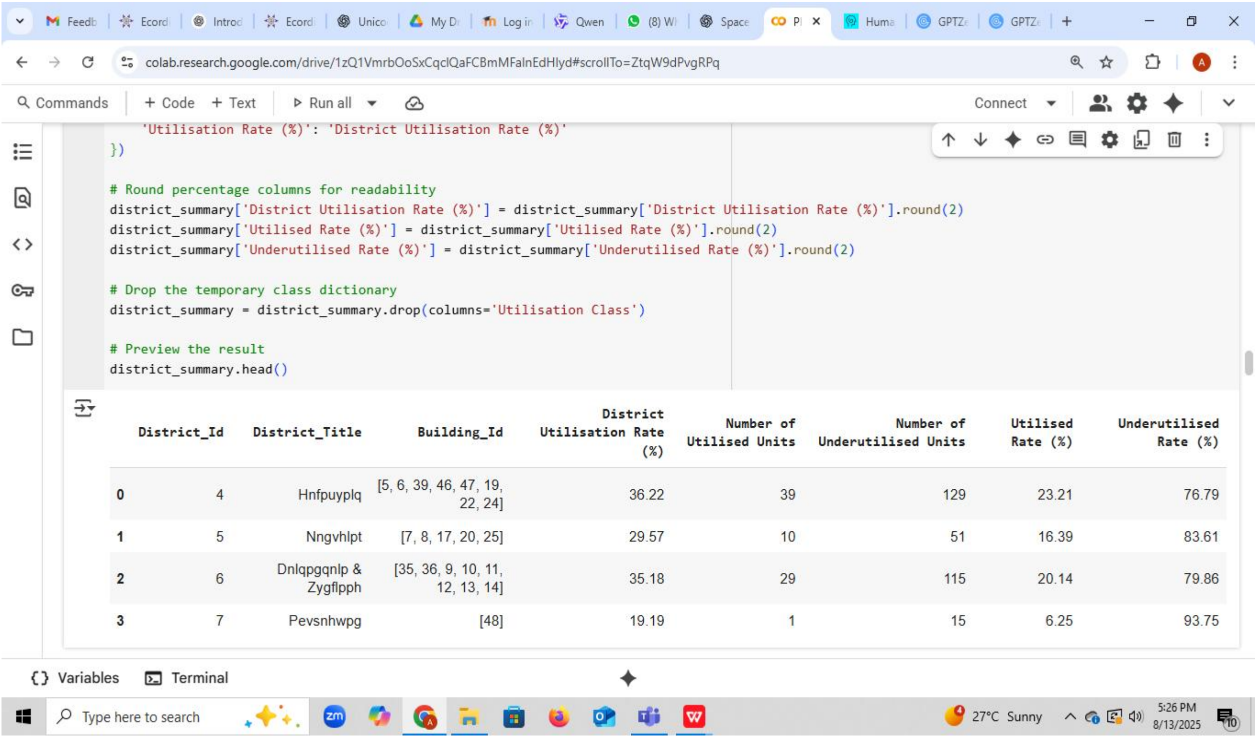


Fig. 24.2: District utilisation rate table_2

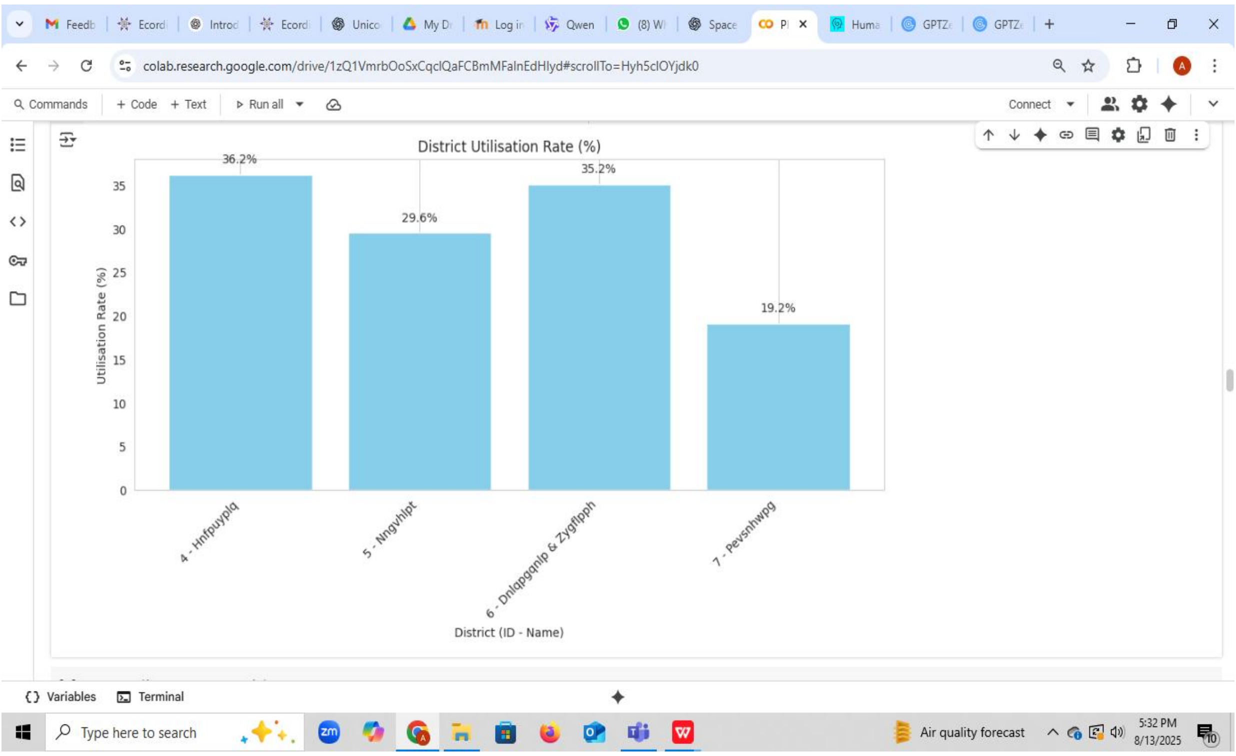


Fig. 24.4: District utilisation rate bar chart_2

Utilised vs underutilised bookable units by district

Fig. 25.2 below shows the graphical distribution of the average utilised and underutilised bookable unit rate by district, comparing them side by side. It can be seen that district id 4 has the highest average utilised bookable rate of 23.20% and the lowest average underutilised bookable unit rate of 76.80% while district id 7 has the lowest utilised average bookable unit rate of 6.20% and the highest average underutilised bookable unit rate of 93.80%. See full code in appendix C (fig. 25.1).

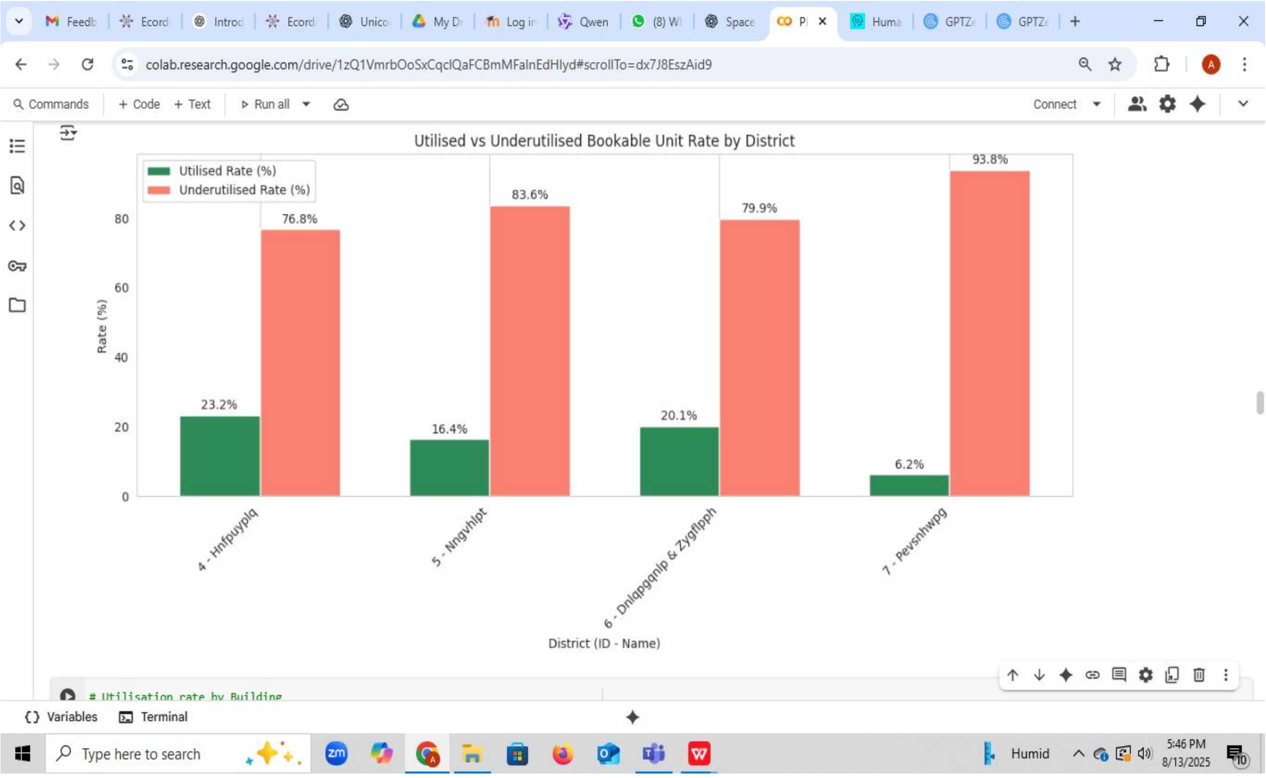


Fig. 25.2: Utilised vs underutilised bookable units by district_2

Building utilisation rate

The building utilisation rate shows the average usage rate of all the bookable units in the building as seen in fig. 26.2. From the visual analysis, building id 17 has the highest utilisation rate of 58.80% while building id 35 has the lowest utilisation rate of 9.40%. See full code in appendix C (fig. 26.1).

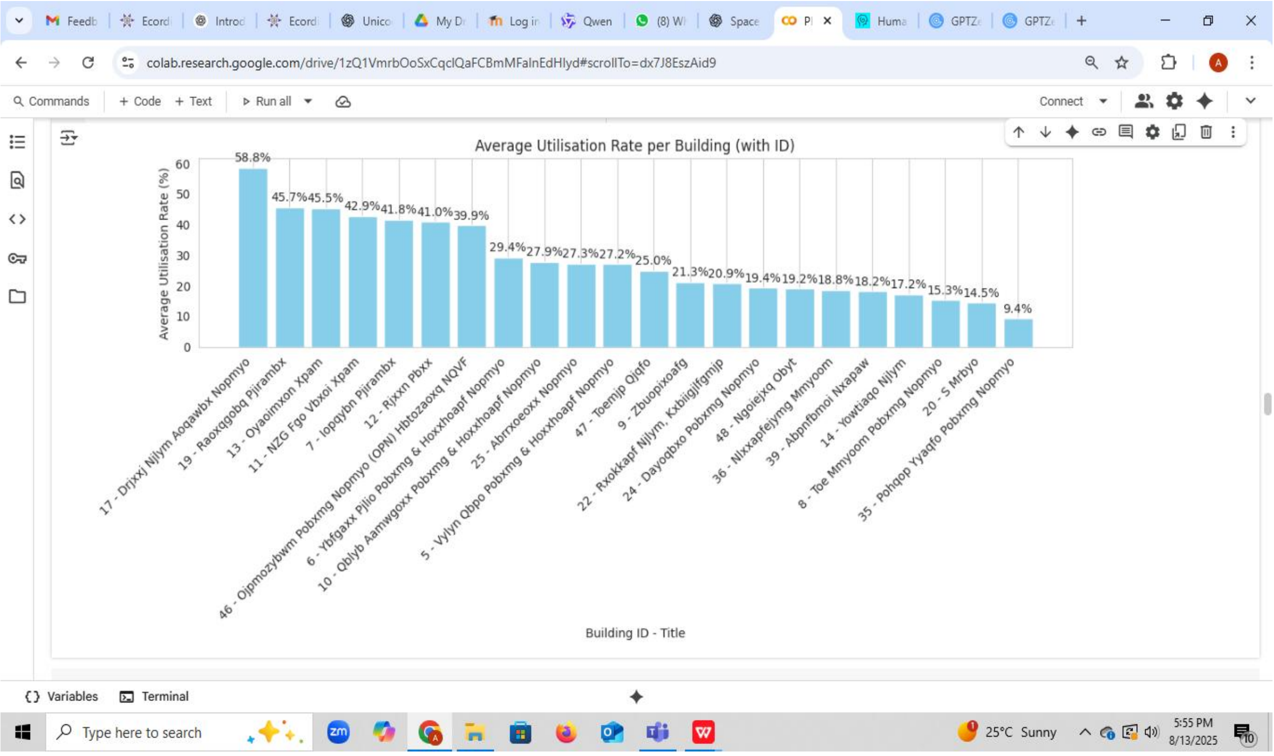


Fig. 26.2: Building utilisation rate_2

Top 10 most utilised bookable units

Fig. 27 below shows the top 10 most utilised bookable units

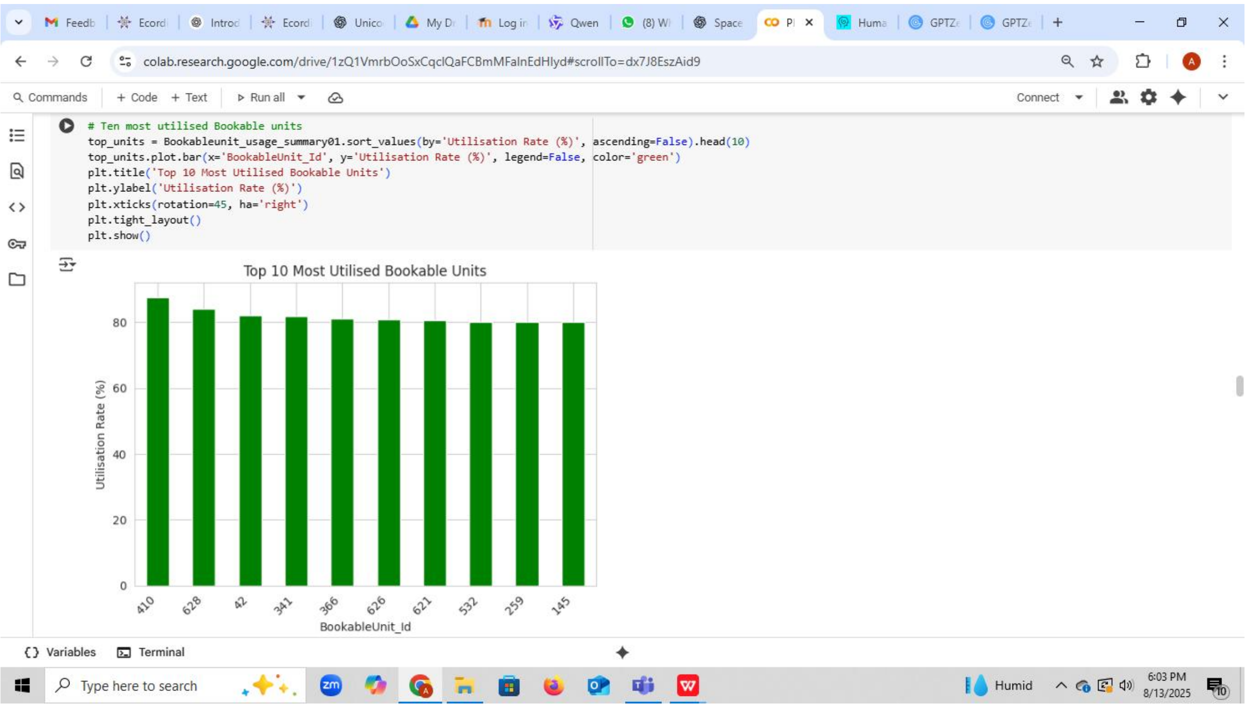


Fig. 27: Top 10 most utilised bookable units

10 least underutilised bookable units

Fig. 28.2 below shows the 10 least underutilised bookable units. See full code in appendix C (fig.28.1).

Fig. 28.2: 10 least underutilised bookable units_2

4.1.3.3 Correlation analysis

The integration of the utilisation rate and yearly booking count datasets as seen in fig. 32 below give a broader horizon of the booking information. It is key to understand how these two approaches correlate and influence the booking system in order to have better insight and foresight of the booking patterns as an overview of the utilisation booking summary shows that some bookable units and buildings have good utilisation rate and low yearly booking count and vice versa. To evaluate this, a correlation analysis is carried out as shown in fig. 29.2 and a heatmap overview of the whole interrelated correlation coefficients as seen in fig. 30.2 below; The correlation analysis output shows that the utilisation rate has weak negative correlation with nine (9) yearly total booking and very weak positive correlation with only two (2) yearly total booking with the heatmap capturing all the correlations among all the yearly total bookings and utilisation rate. This shows that satisfying the usage threshold from historical data does not guarantee that a bookable unit or building is still fit for the booking system operation, thereby making the two approaches (usage and booking count evaluation) imperative in the NHS booking system analysis. See full code in appendix D (fig. 29.1 and fig. 30.1).

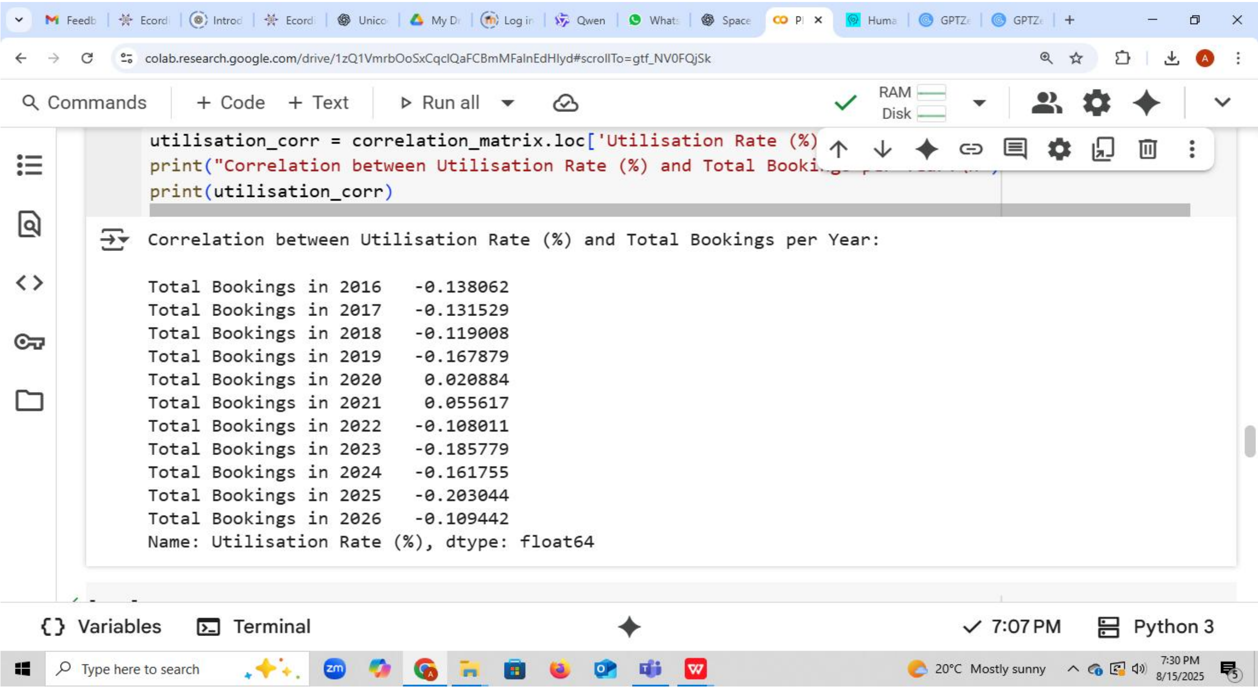


Fig. 29.2: Correlation analysis between utilisation rate and total yearly booking_2

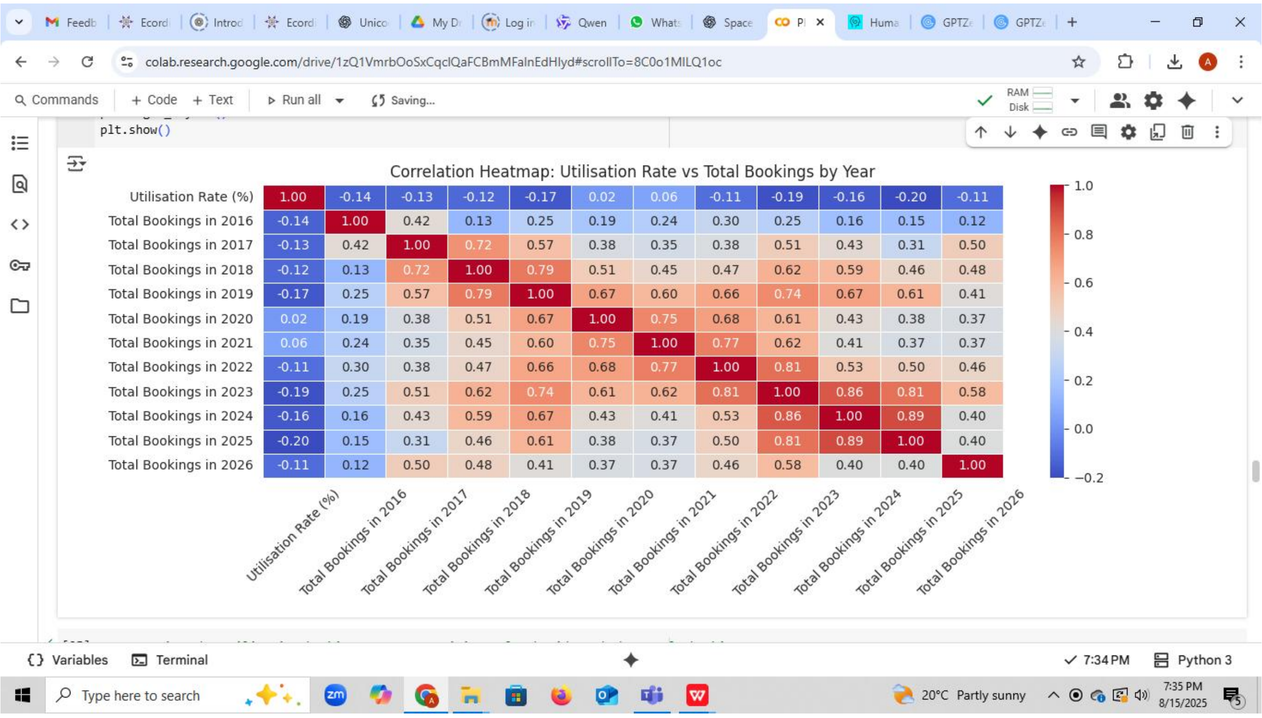


Fig. 30.2: Heatmap plot_2

4.1.3.4 Utilisation booking summary

The space utilisation analysis follows a two-dimensional approach that focus on the usage and booking count analysis for holistic evaluation. In order to have a robust overview of the data, the two approaches are integrated by merging the bookable usage summary data and the yearly booking data to give a consolidated dataset called utilisation booking summary. This dataset shows the utilisation rate for each bookable unit alongside the yearly booking history from 2016 to 2026 (future booking) in a single table as shown in fig. 31.1, fig. 31.2 and fig. 31.3

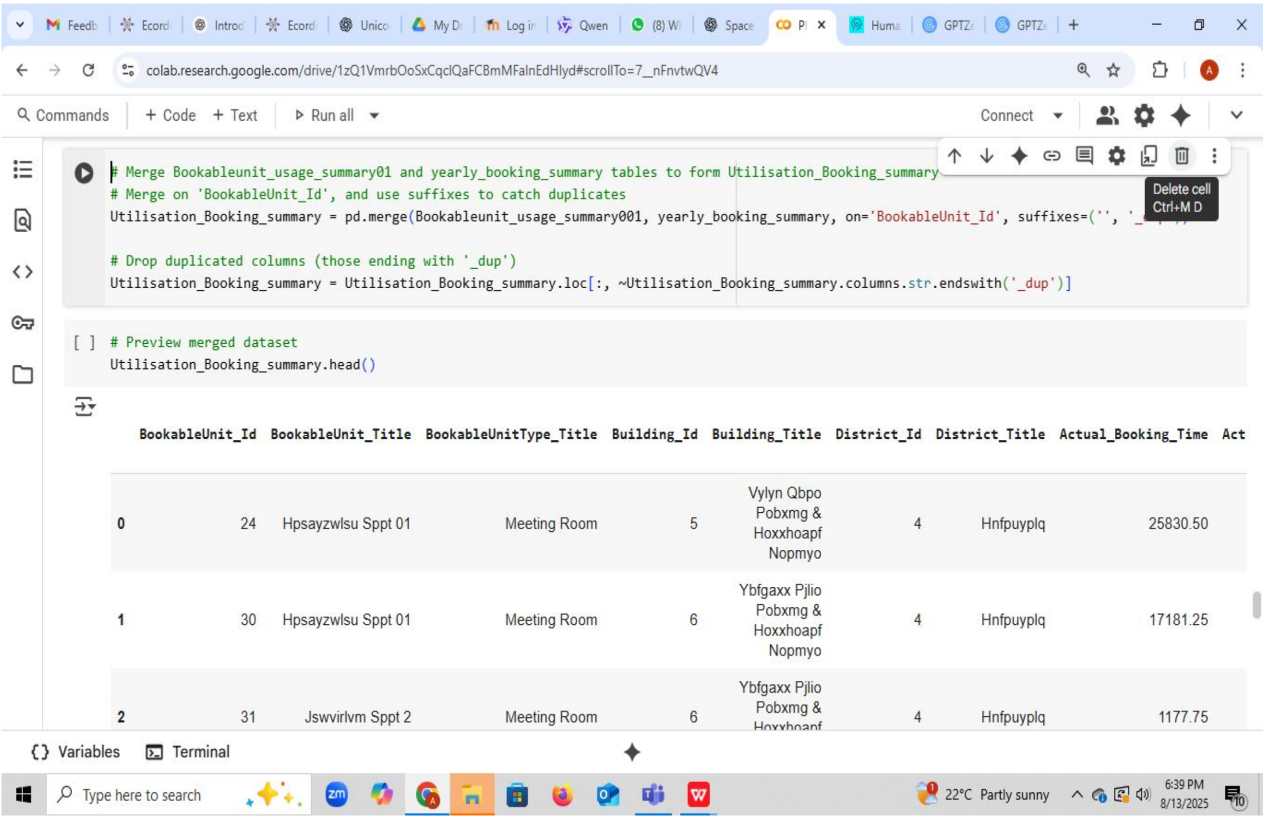
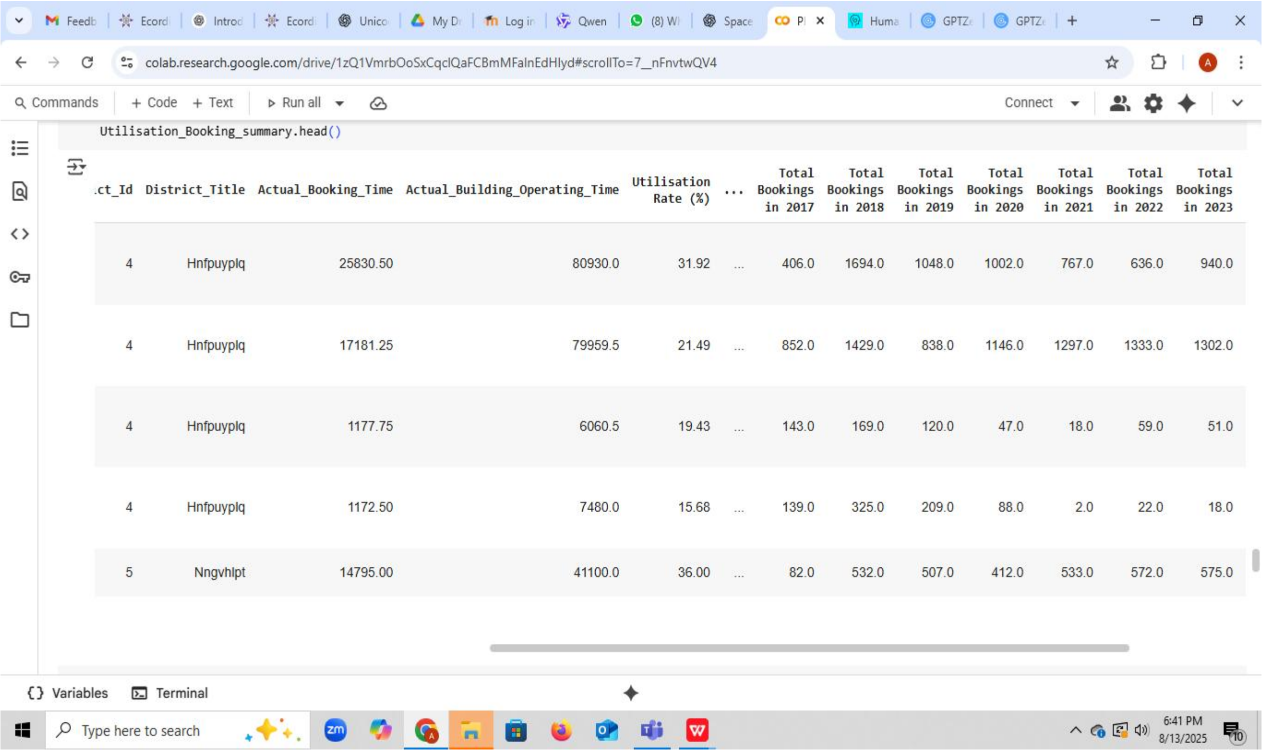


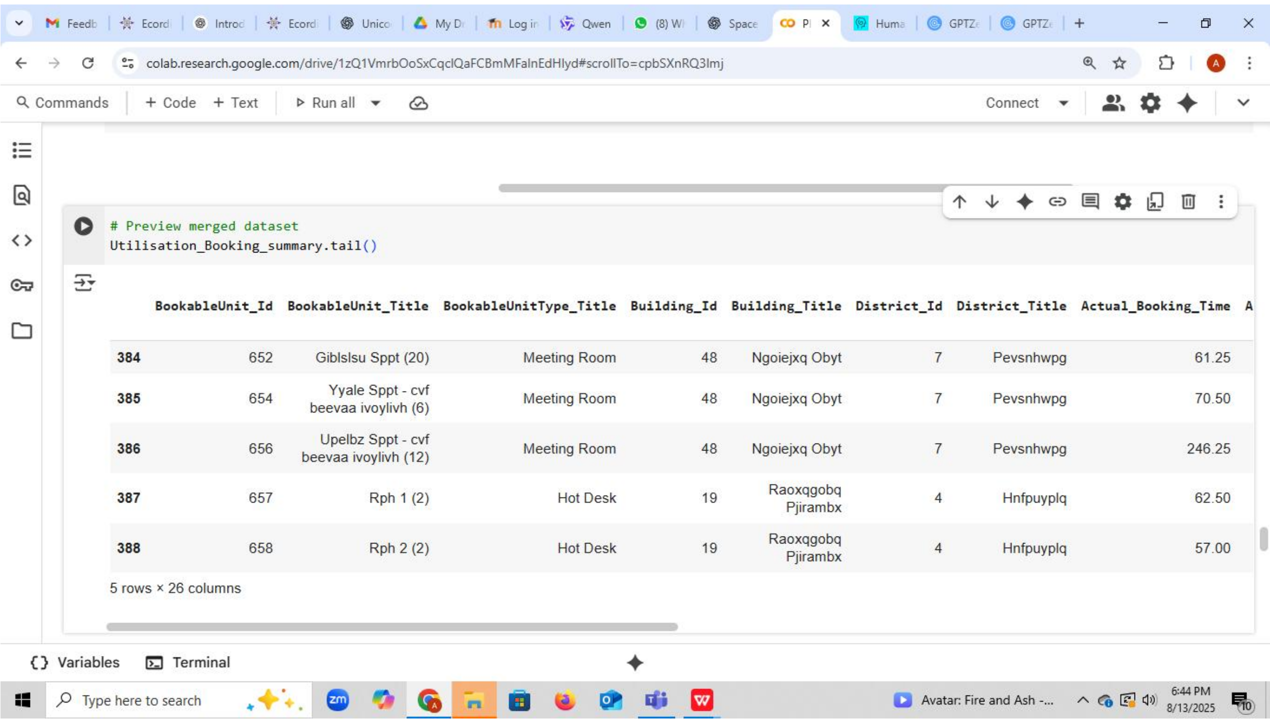
Fig. 31.1: Utilisation booking summary_1



```
Utilisation_Booking_summary.head()
```

...ct_Id	District_Title	Actual_Booking_Time	Actual_Building_Operating_Time	Utilisation Rate (%)	...	Total Bookings in 2017	Total Bookings in 2018	Total Bookings in 2019	Total Bookings in 2020	Total Bookings in 2021	Total Bookings in 2022	Total Bookings in 2023
4	Hnfpuyplq	25830.50	80930.0	31.92	...	406.0	1694.0	1048.0	1002.0	767.0	636.0	940.0
4	Hnfpuyplq	17181.25	79959.5	21.49	...	852.0	1429.0	838.0	1146.0	1297.0	1333.0	1302.0
4	Hnfpuyplq	1177.75	6060.5	19.43	...	143.0	169.0	120.0	47.0	18.0	59.0	51.0
4	Hnfpuyplq	1172.50	7480.0	15.68	...	139.0	325.0	209.0	88.0	2.0	22.0	18.0
5	Nngvhlpt	14795.00	41100.0	36.00	...	82.0	532.0	507.0	412.0	533.0	572.0	575.0

Fig. 31.2: Utilisation booking summary_2



```
# Preview merged dataset
Utilisation_Booking_summary.tail()
```

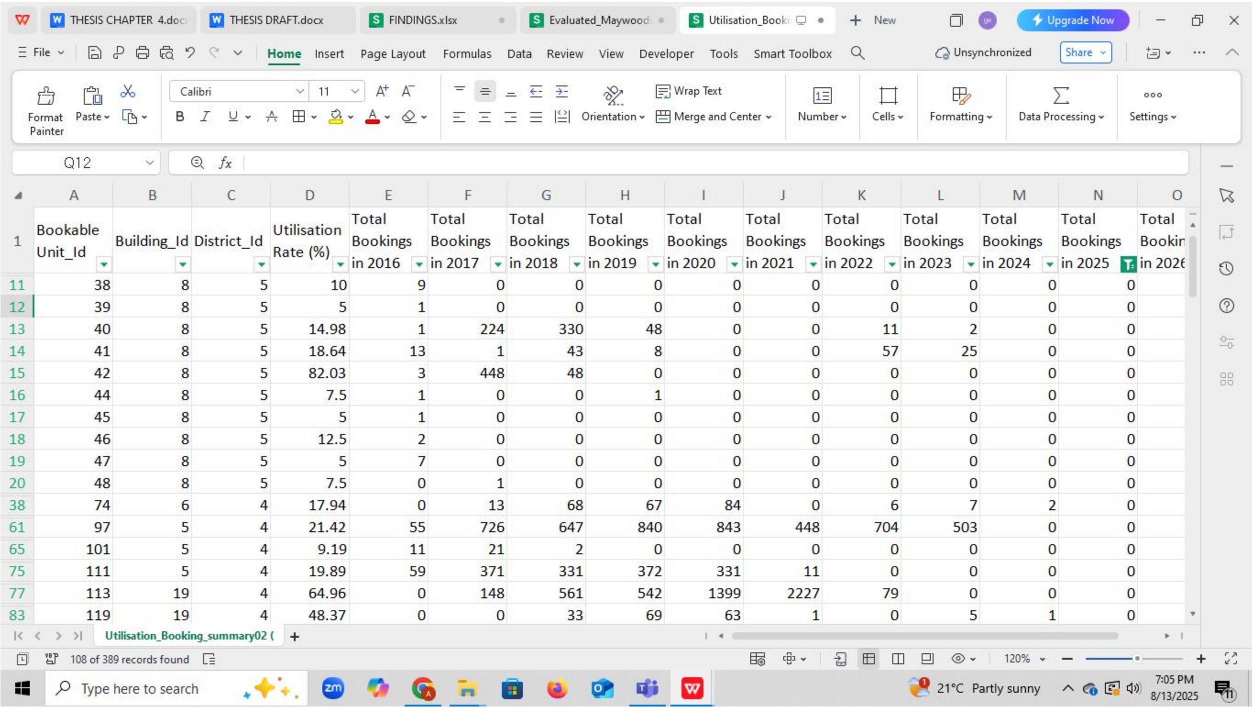
	BookableUnit_Id	BookableUnit_Title	BookableUnitType_Title	Building_Id	Building_Title	District_Id	District_Title	Actual_Booking_Time
384	652	Giblslsu Sppt (20)	Meeting Room	48	Ngoiejxq Obyt	7	Pevsnhwpg	61.25
385	654	Yyale Sppt - cvf beevaa ivoylivh (6)	Meeting Room	48	Ngoiejxq Obyt	7	Pevsnhwpg	70.50
386	656	Upelbz Sppt - cvf beevaa ivoylivh (12)	Meeting Room	48	Ngoiejxq Obyt	7	Pevsnhwpg	246.25
387	657	Rph 1 (2)	Hot Desk	19	Raoxqgobq Pjirambx	4	Hnfpuyplq	62.50
388	658	Rph 2 (2)	Hot Desk	19	Raoxqgobq Pjirambx	4	Hnfpuyplq	57.00

5 rows × 26 columns

Fig. 31.3: Utilisation booking summary_3

Summary of utilisation booking findings

The utilisation booking summary dataset (generated by integrating the yearly booking and bookable usage summary datasets) was exported as a csv file for thorough evaluation to identify patterns in the booking history and usage and from the findings, it can be seen that one hundred and nine (109) bookable units (excluding those dropped during data cleaning and preprocessing such as the ones in district id 3 that were only active in 2016 and those cancelled all through in previous years without recording any active booking) had inconsistent booking patterns and not up to date from historical data as they have been inactive from different previous years down to 2025 and thereby no longer relevant in the booking system. A summary of each bookable unit and building no longer relevant in the booking system is outlined in the redundant buildings and bookable units tables (table 6 and table 7) with the corresponding reasons. A screenshot of the utilisation booking summary csv is shown below in fig. 32.



	Bookable Unit_Id	Building_Id	District_Id	Utilisation Rate (%)	Total Bookings in 2016	Total Bookings in 2017	Total Bookings in 2018	Total Bookings in 2019	Total Bookings in 2020	Total Bookings in 2021	Total Bookings in 2022	Total Bookings in 2023	Total Bookings in 2024	Total Bookings in 2025	Total Bookin in 2026
11	38	8	5	10	9	0	0	0	0	0	0	0	0	0	
12	39	8	5	5	1	0	0	0	0	0	0	0	0	0	
13	40	8	5	14.98	1	224	330	48	0	0	11	2	0	0	
14	41	8	5	18.64	13	1	43	8	0	0	57	25	0	0	
15	42	8	5	82.03	3	448	48	0	0	0	0	0	0	0	
16	44	8	5	7.5	1	0	0	1	0	0	0	0	0	0	
17	45	8	5	5	1	0	0	0	0	0	0	0	0	0	
18	46	8	5	12.5	2	0	0	0	0	0	0	0	0	0	
19	47	8	5	5	7	0	0	0	0	0	0	0	0	0	
20	48	8	5	7.5	0	1	0	0	0	0	0	0	0	0	
38	74	6	4	17.94	0	13	68	67	84	0	6	7	2	0	
61	97	5	4	21.42	55	726	647	840	843	448	704	503	0	0	
65	101	5	4	9.19	11	21	2	0	0	0	0	0	0	0	
75	111	5	4	19.89	59	371	331	372	331	11	0	0	0	0	
77	113	19	4	64.96	0	148	561	542	1399	2227	79	0	0	0	
83	119	19	4	48.37	0	0	33	69	63	1	0	5	1	0	

Fig. 32: Screenshot of utilisation booking summary csv

4.1.3.5 Summary of maywoods dataset after evaluation

After evaluating the utilisation booking dataset to identify configuration with inactive bookings in the years under review (2016 - 2026) and relating it to the maywoods booking dataset (which is the merged dataset before analysis), it was discovered that a total of one (1) district, seven (7) buildings and one hundred and twenty three (123) bookable units have been redundant up till 2025 thereby making them unfit in the operation of the NHS booking system. Below is a tabular and graphical summary of the maywoods dataset after evaluation as seen in table 4, fig. 33, fig. 34.2 and fig. 35.2 showing only active districts, buildings and bookable units, excluding the redundant ones. See full code in appendix E (fig. 34.1 and fig. 35.1).

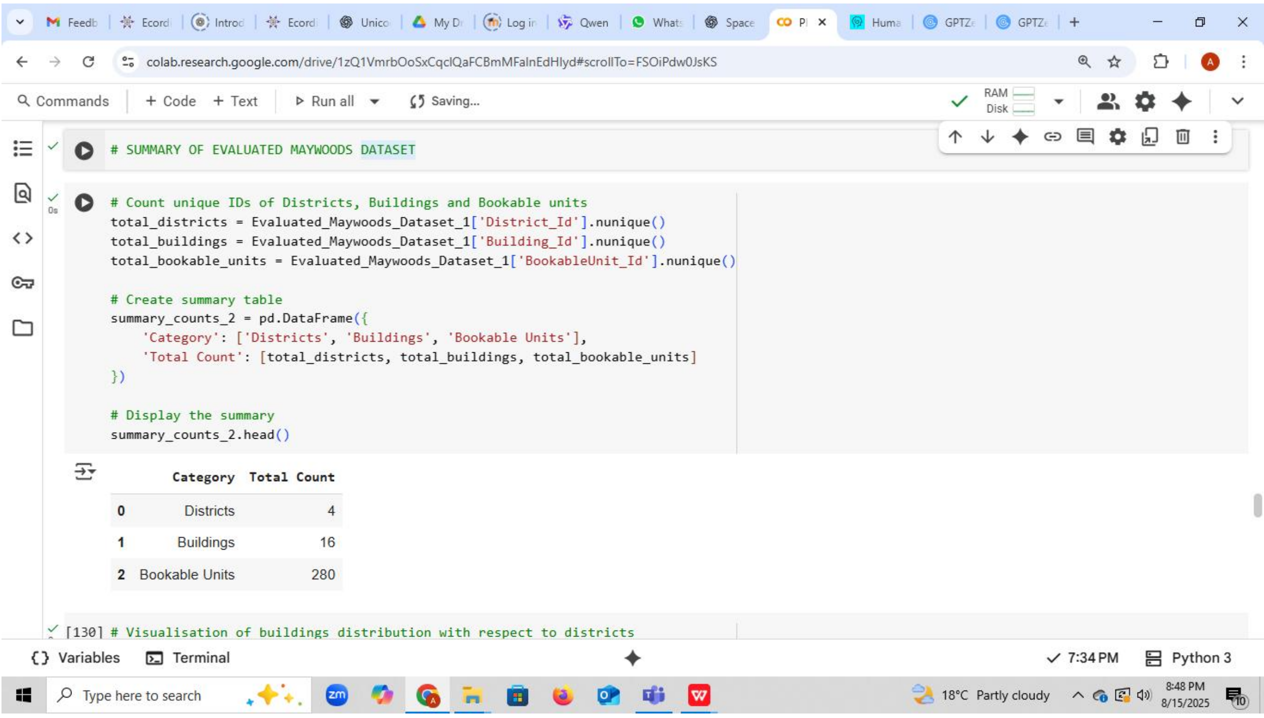


Fig. 33: Maywoods dataset summary after evaluation_1

District_id	Building_id	Bookable unit count
4	5	35

	6	20
	19	39
	22	2
	24	10
	46	6
5	7	17
	8	9
	25	13
6	9	11
	10	33
	11	14
	12	50
	14	3
	36	2
7	48	16
Total count **4**	**16**	**280**

Table 4: Maywoods dataset after evaluation_2

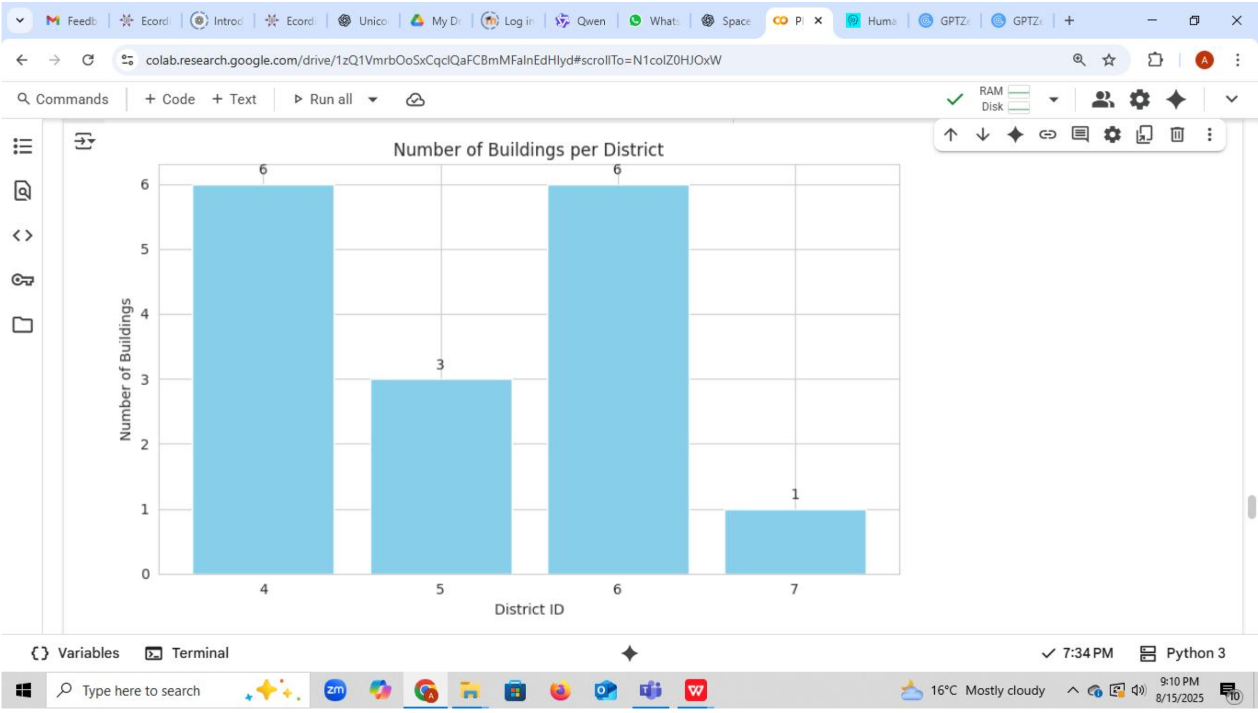


Fig. 34.2: District bar chart after evaluation_2

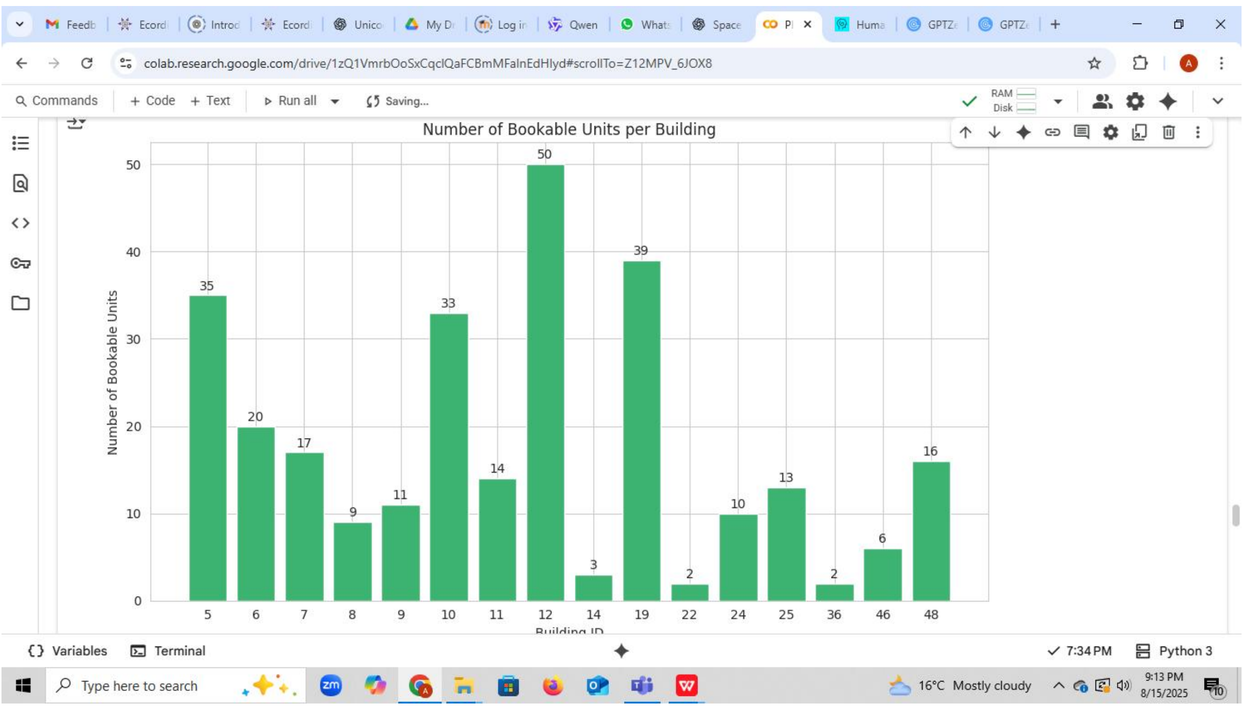


Fig. 35.2: Building bar chart after evaluation_2

Evaluated maywoods dataset

The evaluated maywoods dataset was gotten by dropping the redundant buildings and bookable units (as outlined in table 6 and table 7) from the utilisation booking dataset to retain only the active ones as seen in fig. 36.2. See full code in appendix E (fig. 36.1).

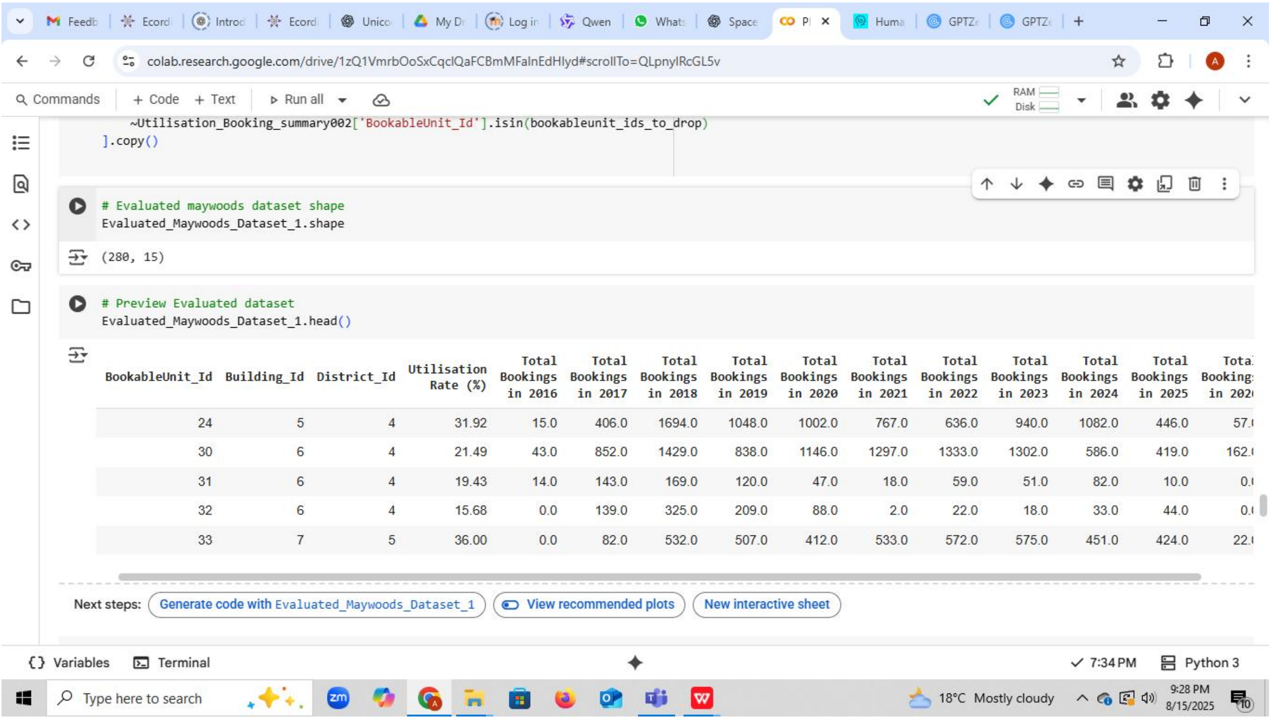


Fig. 36.2: Evaluated maywoods dataset_2

4.1.3.6 Comparison of maywoods dataset before and after evaluation

Table 5, fig. 37.2 and fig. 38.2 below show the tabular and graphical comparison of the maywoods booking dataset before and after evaluation with respect to the districts, buildings and bookable units

District_id	Building_id	Bookable unit count before evaluation	Bookable unit count after evaluation
3	4	4	0

4	5	40	35
	6	21	20
	19	84	39
	22	3	2
	24	11	10
	39	4	0
	46	6	6
	47	1	0
5	7	25	17
	8	19	9
	17	1	0
	20	1	0
	25	17	13
6	9	12	11
	10	45	33
	11	20	14
	12	55	50
	13	10	0
	14	3	3

	35	1	0
	36	3	2
7	48	17	16
TOTAL		**403**	**280**

Table 5: Tabular comparison of maywoods dataset before and after evaluation

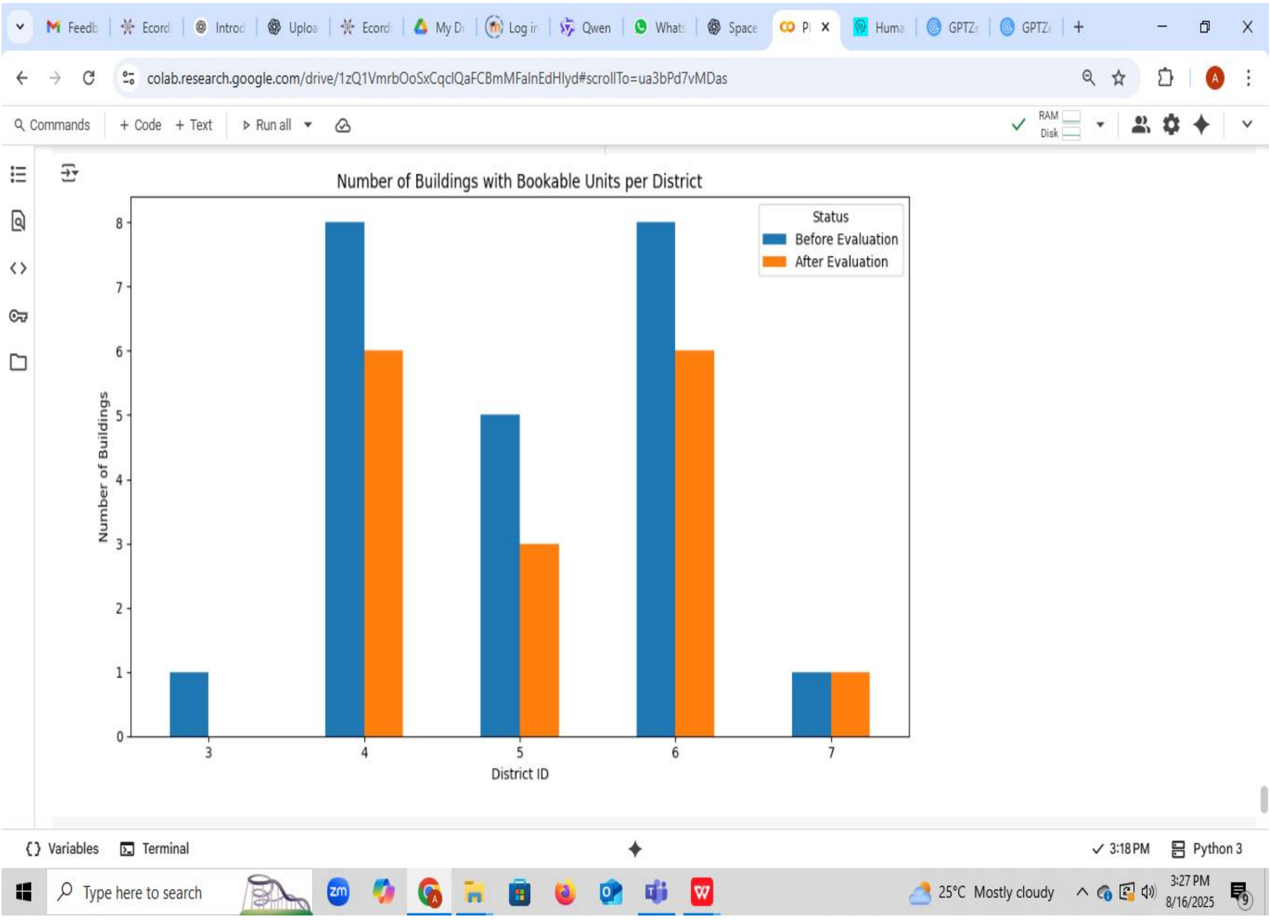


Fig. 37.2: Comparison of districts before and after evaluation_2

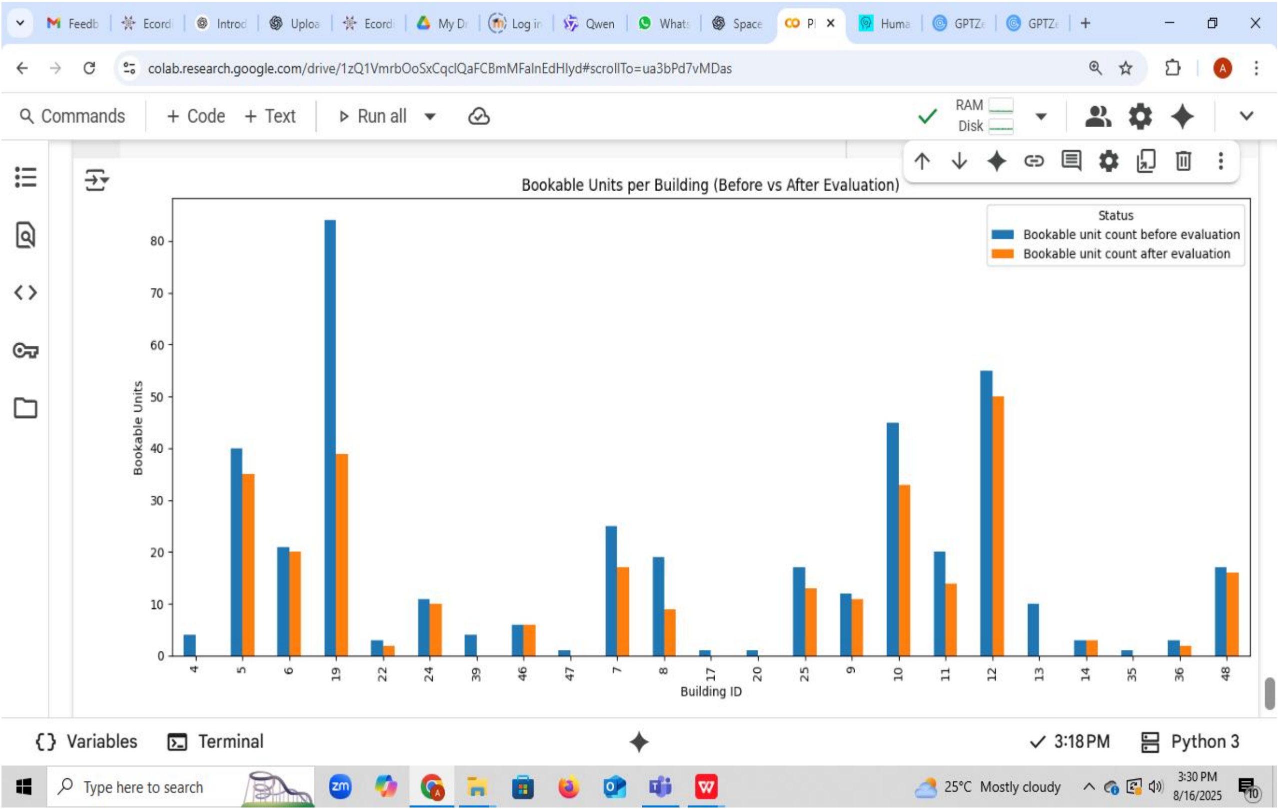
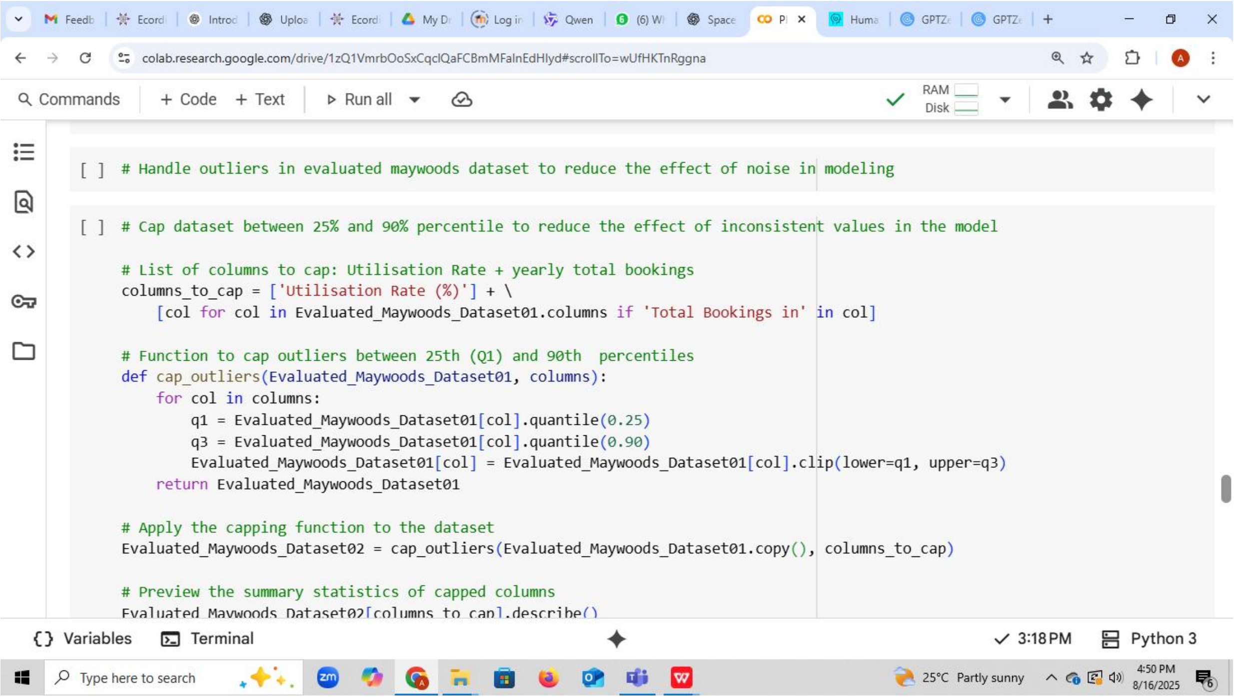


Fig. 38.2: Comparison of buildings before and after evaluation_2

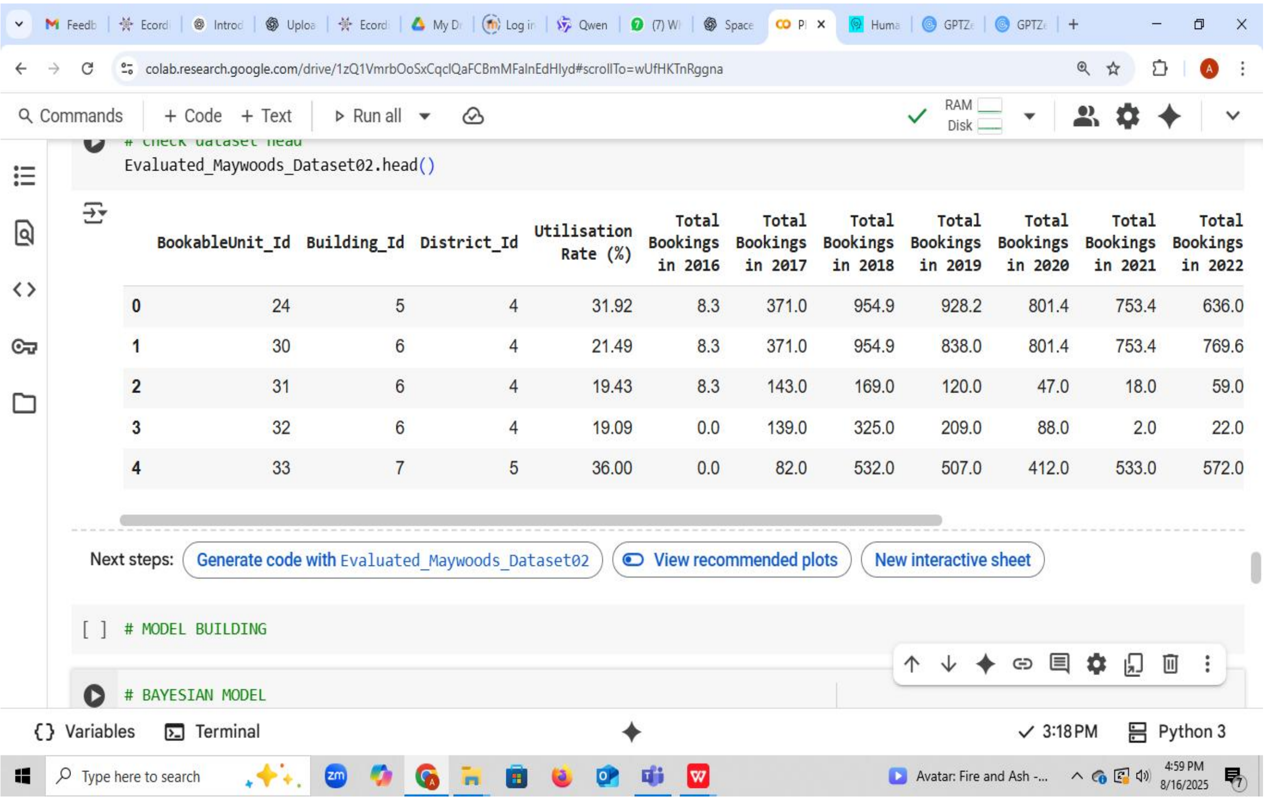
4.1.4 Modeling

In this study, five (5) predictive models namely; Bayesian ridge regression model, Linear regression model, Random forest model, Gradient boosting regressor model and Linear support vector regression model are built and evaluated using the mean absolute error in order to get the best model to forecast the total yearly booking for 2026. Before modeling, feature engineering was carried out on the evaluated maywoods dataset by handling possible outliers with 25% lower boundary capping and 90% upper boundary capping as seen in fig. 39 and excluding some columns (2016, 2017, 2025 and 2026) in order to reduce the impact of noise in the model building process. 2016 and 2017 have too much noise in their data while 2025 have incomplete data since it captured bookings from January to May and 2026 contains incomplete future bookings. Below is the summary of the formatted dataset as seen in fig. 40.



```
[ ] # Handle outliers in evaluated maywoods dataset to reduce the effect of noise in modeling

[ ] # Cap dataset between 25% and 90% percentile to reduce the effect of inconsistent values in the model

    # List of columns to cap: Utilisation Rate + yearly total bookings
    columns_to_cap = ['Utilisation Rate (%)'] + \
        [col for col in Evaluated_Maywoods_Dataset01.columns if 'Total Bookings in' in col]

    # Function to cap outliers between 25th (Q1) and 90th  percentiles
    def cap_outliers(Evaluated_Maywoods_Dataset01, columns):
        for col in columns:
            q1 = Evaluated_Maywoods_Dataset01[col].quantile(0.25)
            q3 = Evaluated_Maywoods_Dataset01[col].quantile(0.90)
            Evaluated_Maywoods_Dataset01[col] = Evaluated_Maywoods_Dataset01[col].clip(lower=q1, upper=q3)
        return Evaluated_Maywoods_Dataset01

    # Apply the capping function to the dataset
    Evaluated_Maywoods_Dataset02 = cap_outliers(Evaluated_Maywoods_Dataset01.copy(), columns_to_cap)

    # Preview the summary statistics of capped columns
    Evaluated_Maywoods_Dataset02[columns_to_cap].describe()
```

Fig. 39: Handling outliers



```
Evaluated_Maywoods_Dataset02.head()
```

	BookableUnit_Id	Building_Id	District_Id	Utilisation Rate (%)	Total Bookings in 2016	Total Bookings in 2017	Total Bookings in 2018	Total Bookings in 2019	Total Bookings in 2020	Total Bookings in 2021	Total Bookings in 2022
0	24	5	4	31.92	8.3	371.0	954.9	928.2	801.4	753.4	636.0
1	30	6	4	21.49	8.3	371.0	954.9	838.0	801.4	753.4	769.6
2	31	6	4	19.43	8.3	143.0	169.0	120.0	47.0	18.0	59.0
3	32	6	4	19.09	0.0	139.0	325.0	209.0	88.0	2.0	22.0
4	33	7	5	36.00	0.0	82.0	532.0	507.0	412.0	533.0	572.0

```
[ ] # MODEL BUILDING

# BAYESIAN MODEL
```

Fig. 40: Formatted dataset

Bayesian ridge regression model

The Bayesian ridge regression model is a statistical model that is also used as a machine learning model for regression tasks.. This approach produced probabilistic predictions capturing uncertainty around them. One of its unique quality is that it accounts for uncertainty rather than giving just a single prediction. Below is the model training and testing output as seen in fig. 41.3. See full code in appendix F (fig. 41.1 and fig. 41.2).

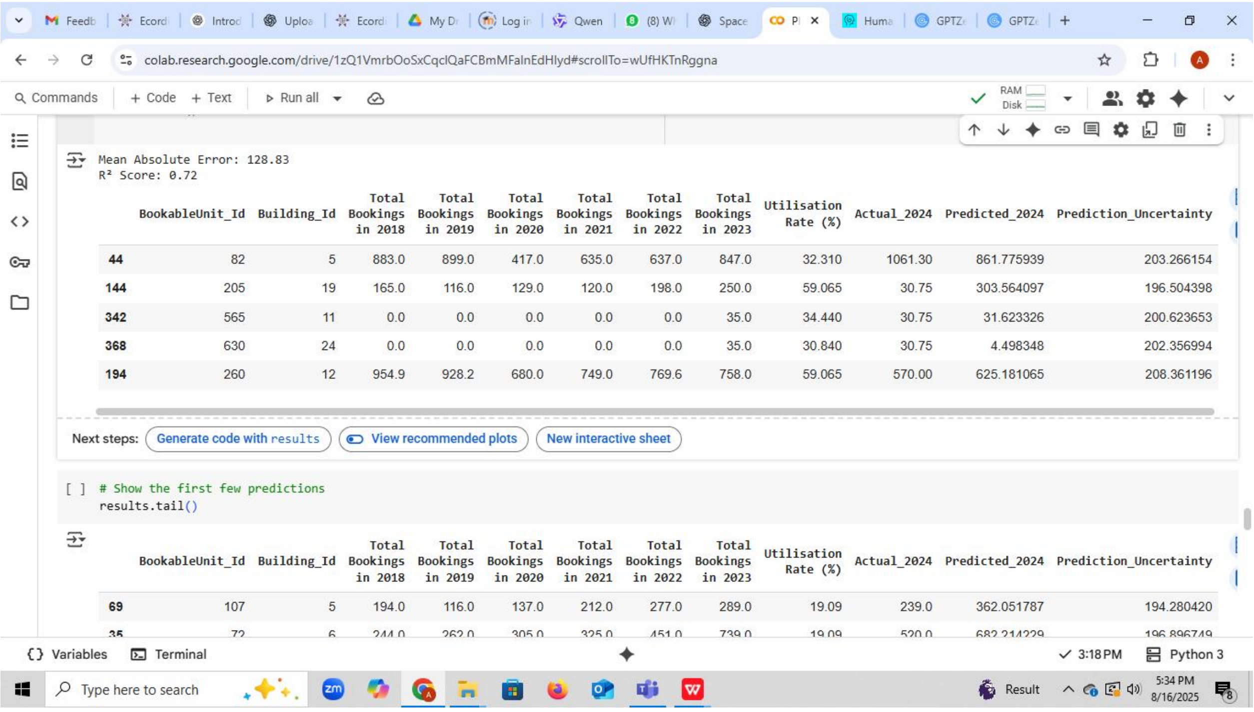
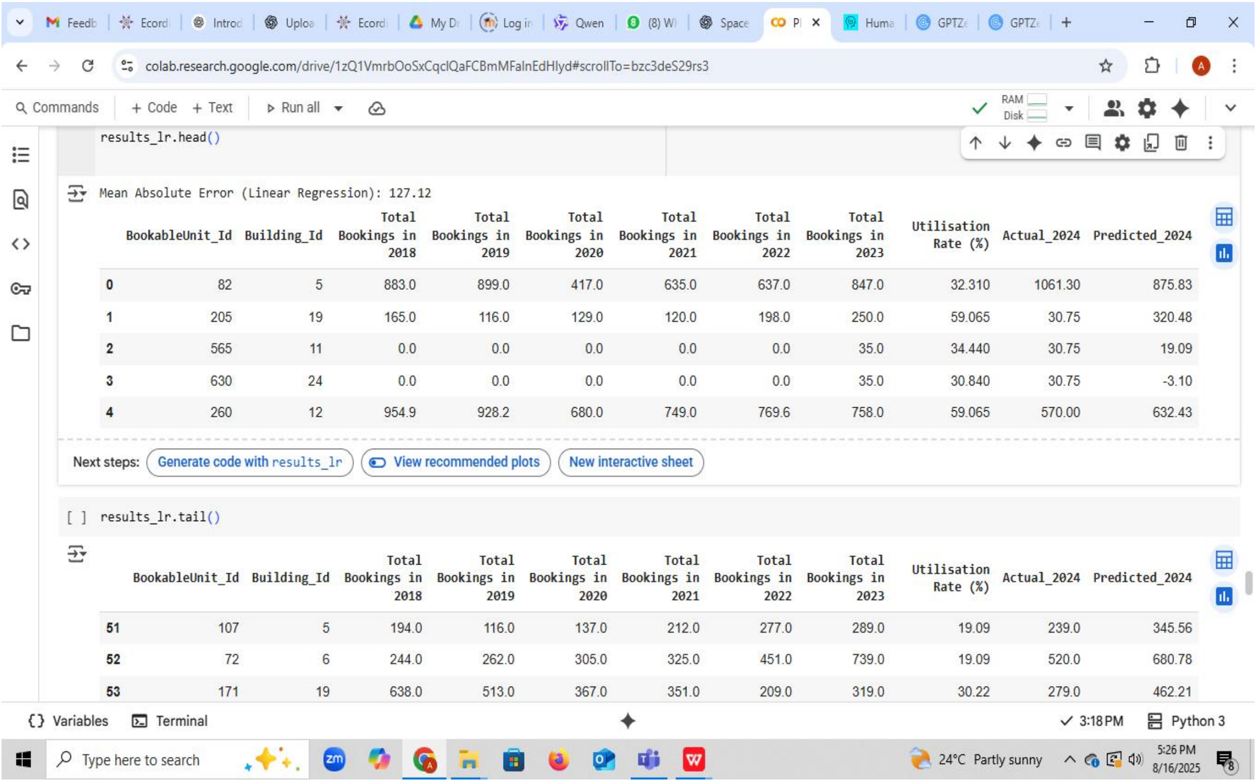


Fig. 41.3: Bayesian ridge regression model testing output

From the outcome of the model testing, the mean absolute error is 128.83 while the R2 (R square) Score is 0.72. This implies that on average, the model's predictions are off by about 129 units from the actual values with the model explaining about 72% variation in the data.

Linear regression model

The linear regression is a statistical model that is also a regression model in machine learning. It is used to explain the relationship of a dependent variable with one or more independent variables. The model predicts an outcome as a straight line relationship between predictors and the response. Fig. 42.3 below shows the model training and testing output. See full code in appendix F (fig. 42.1 and fig. 42.2).



```
results_lr.head()
```

Mean Absolute Error (Linear Regression): 127.12

	BookableUnit_Id	Building_Id	Total Bookings in 2018	Total Bookings in 2019	Total Bookings in 2020	Total Bookings in 2021	Total Bookings in 2022	Total Bookings in 2023	Utilisation Rate (%)	Actual_2024	Predicted_2024
0	82	5	883.0	899.0	417.0	635.0	637.0	847.0	32.310	1061.30	875.83
1	205	19	165.0	116.0	129.0	120.0	198.0	250.0	59.065	30.75	320.48
2	565	11	0.0	0.0	0.0	0.0	0.0	35.0	34.440	30.75	19.09
3	630	24	0.0	0.0	0.0	0.0	0.0	35.0	30.840	30.75	-3.10
4	260	12	954.9	928.2	680.0	749.0	769.6	758.0	59.065	570.00	632.43

Next steps: Generate code with results_lr View recommended plots New interactive sheet

```
[ ] results_lr.tail()
```

	BookableUnit_Id	Building_Id	Total Bookings in 2018	Total Bookings in 2019	Total Bookings in 2020	Total Bookings in 2021	Total Bookings in 2022	Total Bookings in 2023	Utilisation Rate (%)	Actual_2024	Predicted_2024
51	107	5	194.0	116.0	137.0	212.0	277.0	289.0	19.09	239.0	345.56
52	72	6	244.0	262.0	305.0	325.0	451.0	739.0	19.09	520.0	680.78
53	171	19	638.0	513.0	367.0	351.0	209.0	319.0	30.22	279.0	462.21

Fig. 42.3: Linear regression model testing output

The mean absolute error from the output is 127.12. This implies that on average, the Linear Regression model's predictions are off by about 127 units compared to the actual values.

Random forest model

This is a machine learning model that uses the ensemble of decision tree to make predictions. Instead of relying on a single tree which may over fit the data, it builds many trees and combine their results for stronger and more reliable predictions. It handles none linear relationship and complex interactions. Fig. 43.3 below shows the model training and testing output. See full code in appendix F (fig. 43.1 and fig. 43.2).

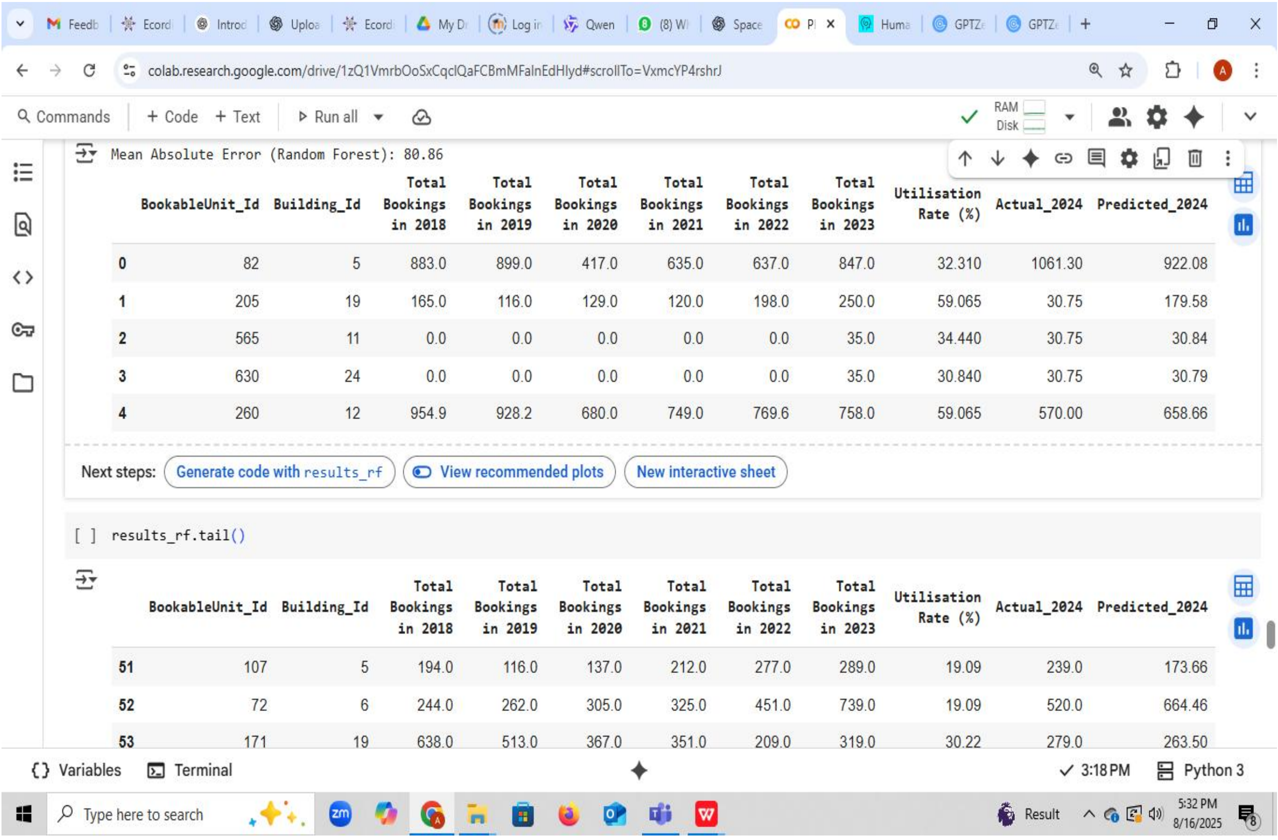
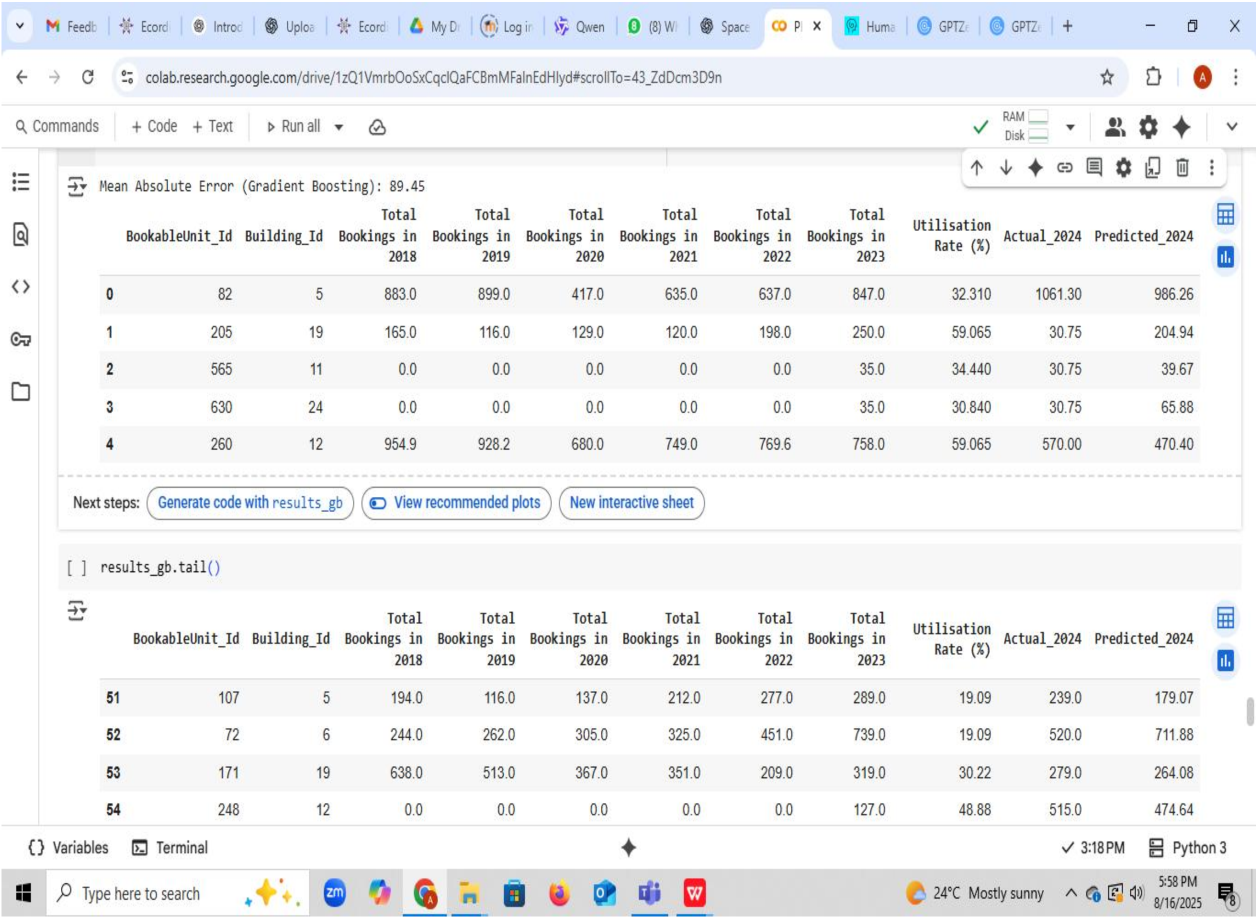


Fig. 43.3: Random forest testing output

The mean absolute error from the output of the model is 80.86. This implies that on average, the Random forest model's predictions are off by about 81 units compared to the actual values.

Gradient boosting regressor

The gradient boosting regressor is a machine learning model that combines many weak learners (usually shallow decision tress) in a sequential way to build predictions. It also captures non- linear relationship and complex interactions. Fig. 44.3 below shows the model training and testing output. See full code in appendix F (fig. 44.1 and fig. 44.2).



Mean Absolute Error (Gradient Boosting): 89.45

	BookableUnit_Id	Building_Id	Total Bookings in 2018	Total Bookings in 2019	Total Bookings in 2020	Total Bookings in 2021	Total Bookings in 2022	Total Bookings in 2023	Utilisation Rate (%)	Actual_2024	Predicted_2024
0	82	5	883.0	899.0	417.0	635.0	637.0	847.0	32.310	1061.30	986.26
1	205	19	165.0	116.0	129.0	120.0	198.0	250.0	59.065	30.75	204.94
2	565	11	0.0	0.0	0.0	0.0	0.0	35.0	34.440	30.75	39.67
3	630	24	0.0	0.0	0.0	0.0	0.0	35.0	30.840	30.75	65.88
4	260	12	954.9	928.2	680.0	749.0	769.6	758.0	59.065	570.00	470.40

```
results_gb.tail()
```

	BookableUnit_Id	Building_Id	Total Bookings in 2018	Total Bookings in 2019	Total Bookings in 2020	Total Bookings in 2021	Total Bookings in 2022	Total Bookings in 2023	Utilisation Rate (%)	Actual_2024	Predicted_2024
51	107	5	194.0	116.0	137.0	212.0	277.0	289.0	19.09	239.0	179.07
52	72	6	244.0	262.0	305.0	325.0	451.0	739.0	19.09	520.0	711.88
53	171	19	638.0	513.0	367.0	351.0	209.0	319.0	30.22	279.0	264.08
54	248	12	0.0	0.0	0.0	0.0	0.0	127.0	48.88	515.0	474.64

Fig. 44.3: Gradient boosting regressor testing output

The mean absolute error from the output of the model is 89.45. This implies that on average, the Gradient boosting regressor model's predictions are off by about 89 units compared to the actual values.

Linear Support Vector Regression model

This is a regression approach that uses the principles of support vector machine. It tries to fit a line that keeps many data points as much as possible within a margin of tolerance. Fig. 45.3 below shows the model training and testing output. See full code in appendix F (fig. 45.1 and fig. 45.2).

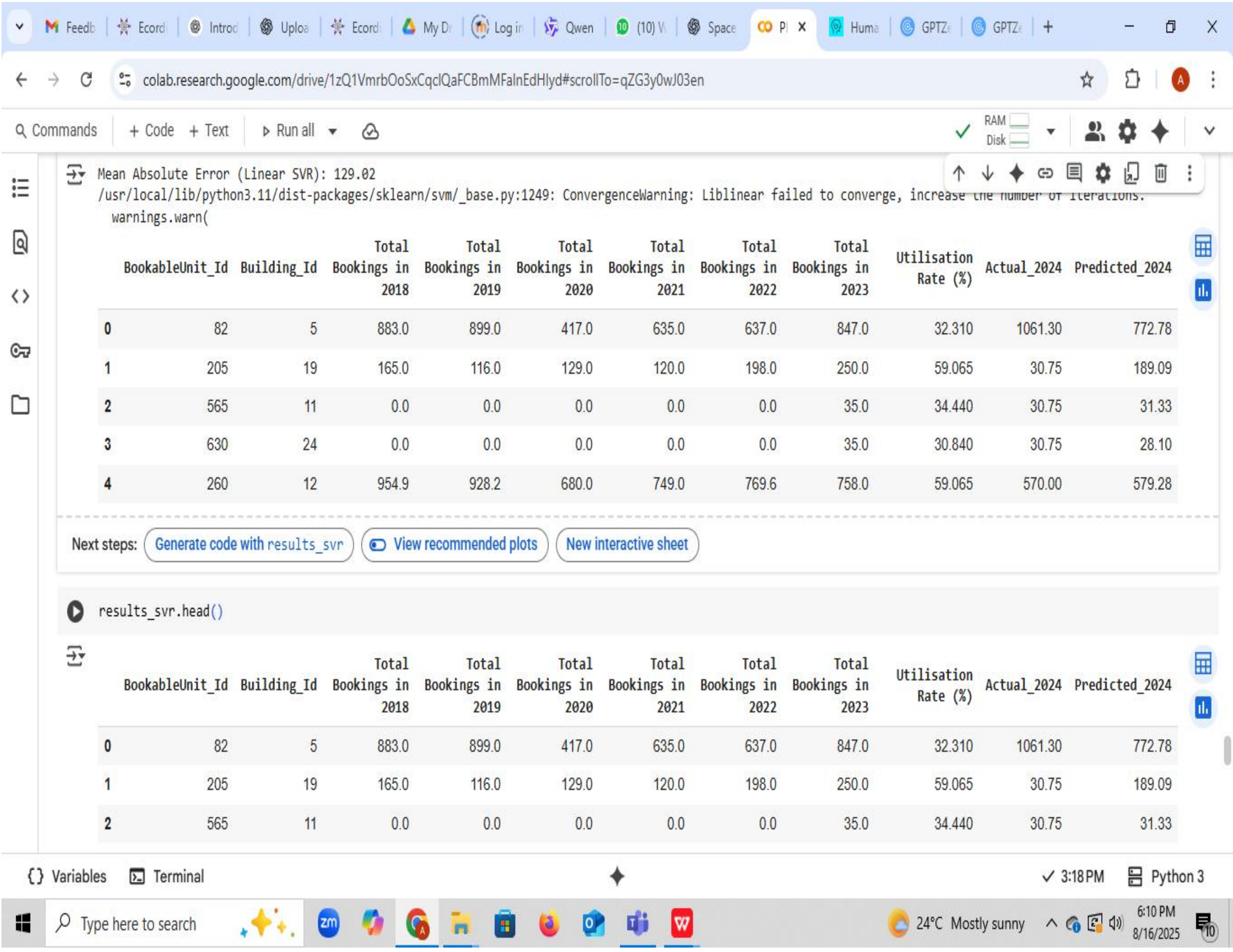


Fig. 45.3: Support vector regression model output

The mean absolute error from the output of the model is 129.02. This implies that on average, the Gradient boosting regressor model's predictions are off by about 129 units compared to the actual values.

4.1.4.1 Model evaluation

When building predictive models, it is important to measure how close the predictions are to the actual values. In this study, this is done using the mean absolute error metric, which helps to compare models and choose the best one. Below is a graphical evaluation of the models error metric as seen in fig. 46.2. See full code in appendix F (fig. 46.1).

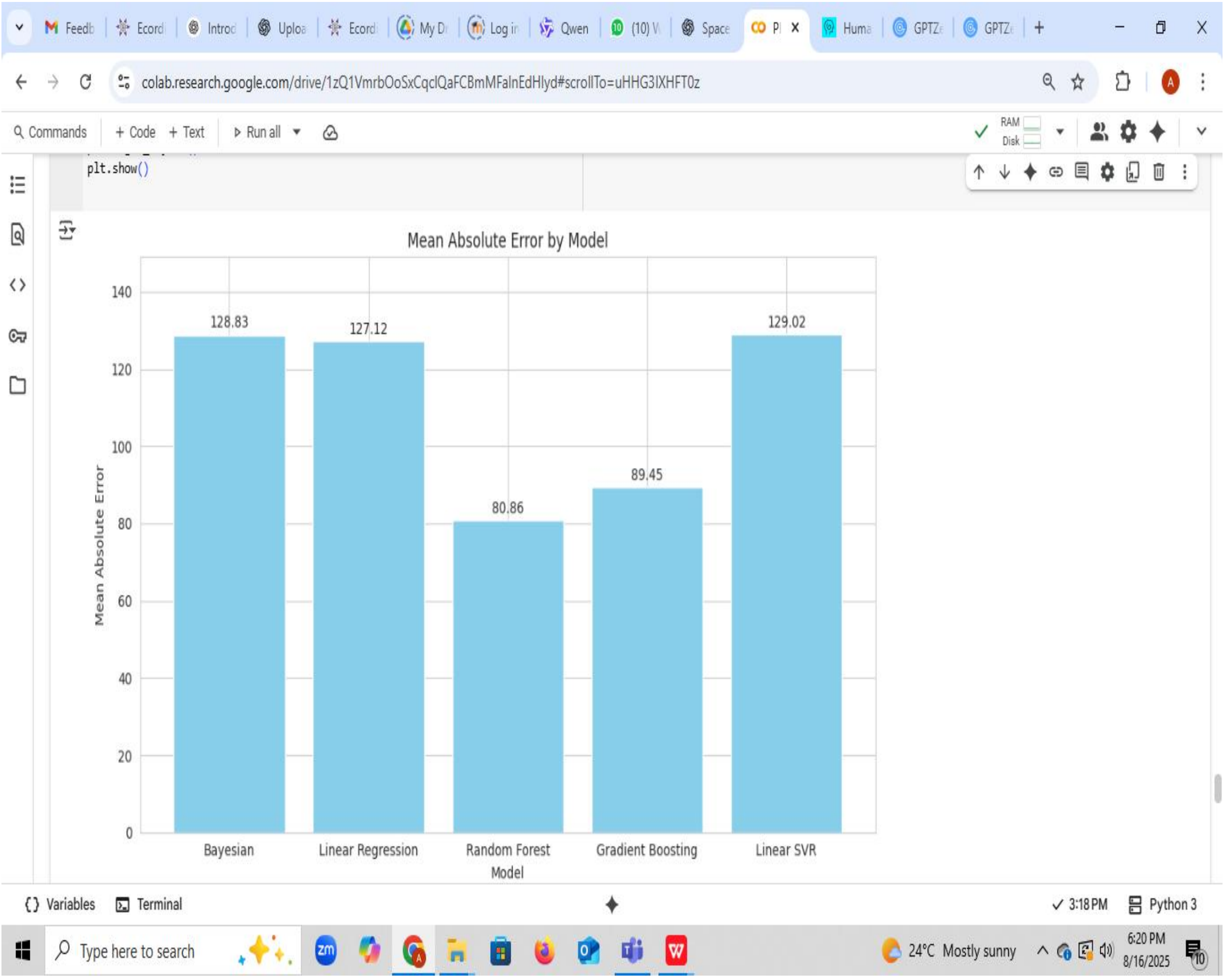


Fig. 46.2: Model evaluation_2

From the chart, it can be seen that random forest model has the least mean absolute error, making it the best fit predictive model and linear support vector regression has the highest mean absolute error thereby making it the worst fit model for forecasting compare to the other models.

Random Forest graphical evaluation

Below is the line plot of the random forest model test, comparing the patterns of the predicted values to the actual values as shown in fig. 47.2. See full code in appendix F (fig. 47.1).

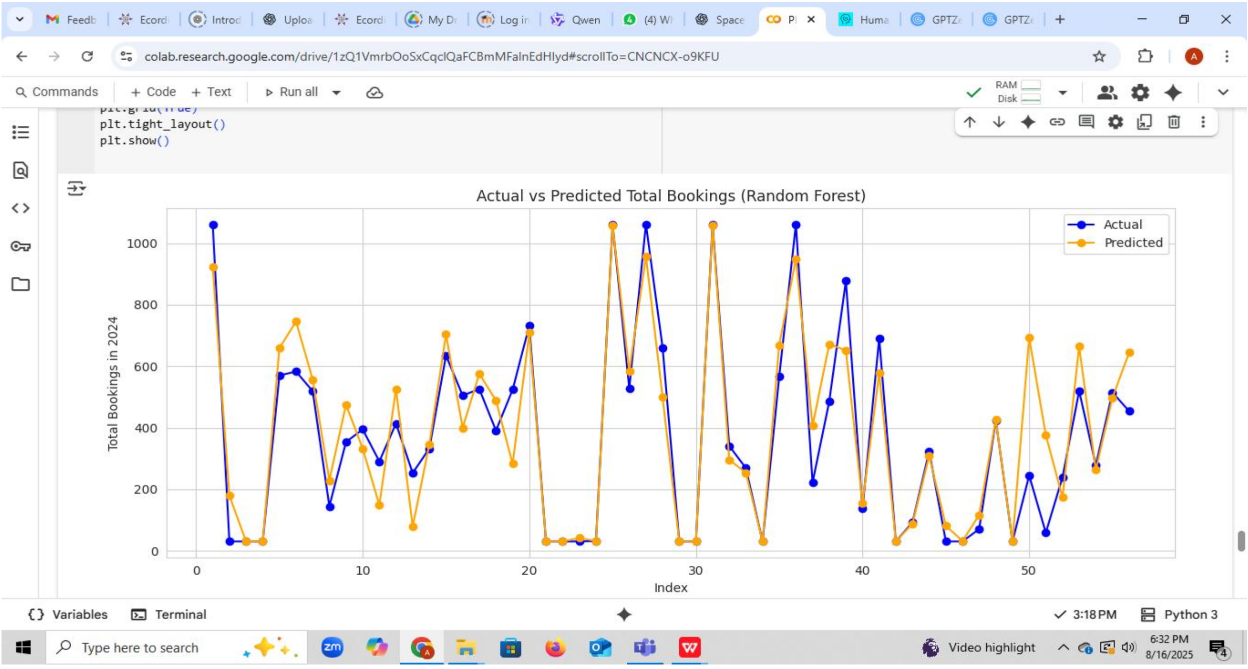


Fig. 47.2: Random forest model evaluation plot

4.1.4.2 Prediction for total booking 2026

To implement the 2026 total booking forecasting, total booking year 2025 was scaled up as it contained data from January to May only. After that, the random forest model is then retrained on the entire evaluated dataset before prediction. Below is the outcome as seen in fig. 48.3. See full code in appendix F (fig. 48.1 and fig. 48.2).

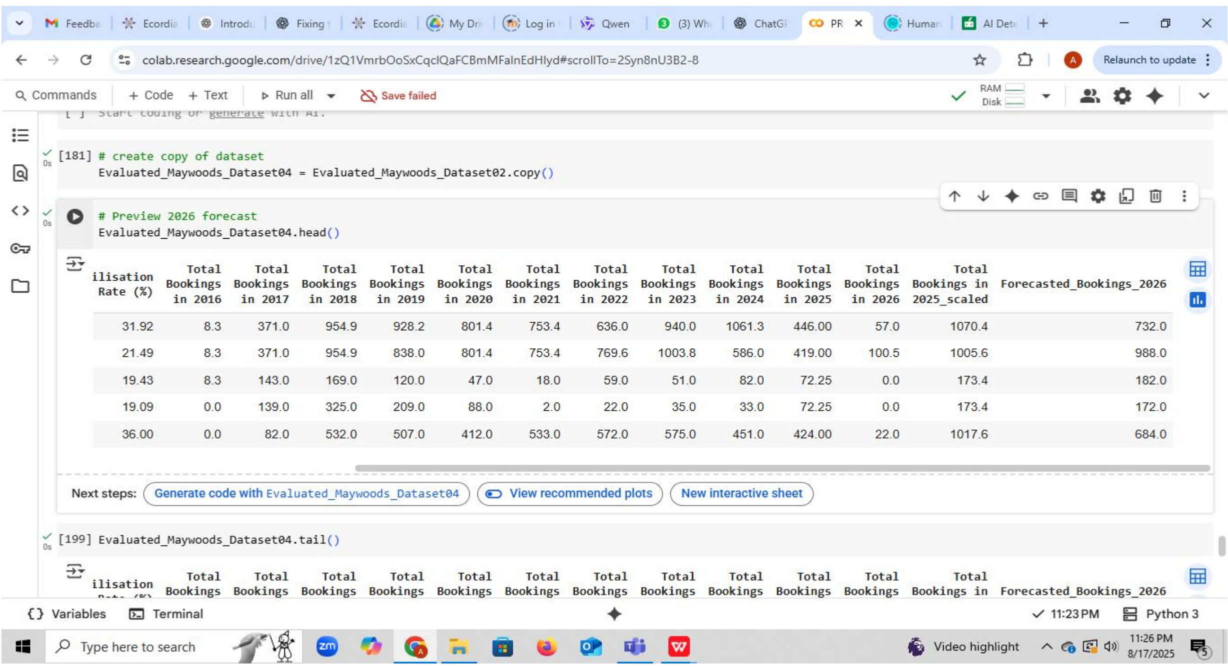


Fig. 48.3: Forecast for total booking 2026 (3)

Preview of yearly booking of evaluated data with 2026 forecast

Below is the yearly booking plot with 2026 forecast inclusive as shown in fig. 49.3. See full code in appendix F (fig. 49.1 and fig. 49.2).

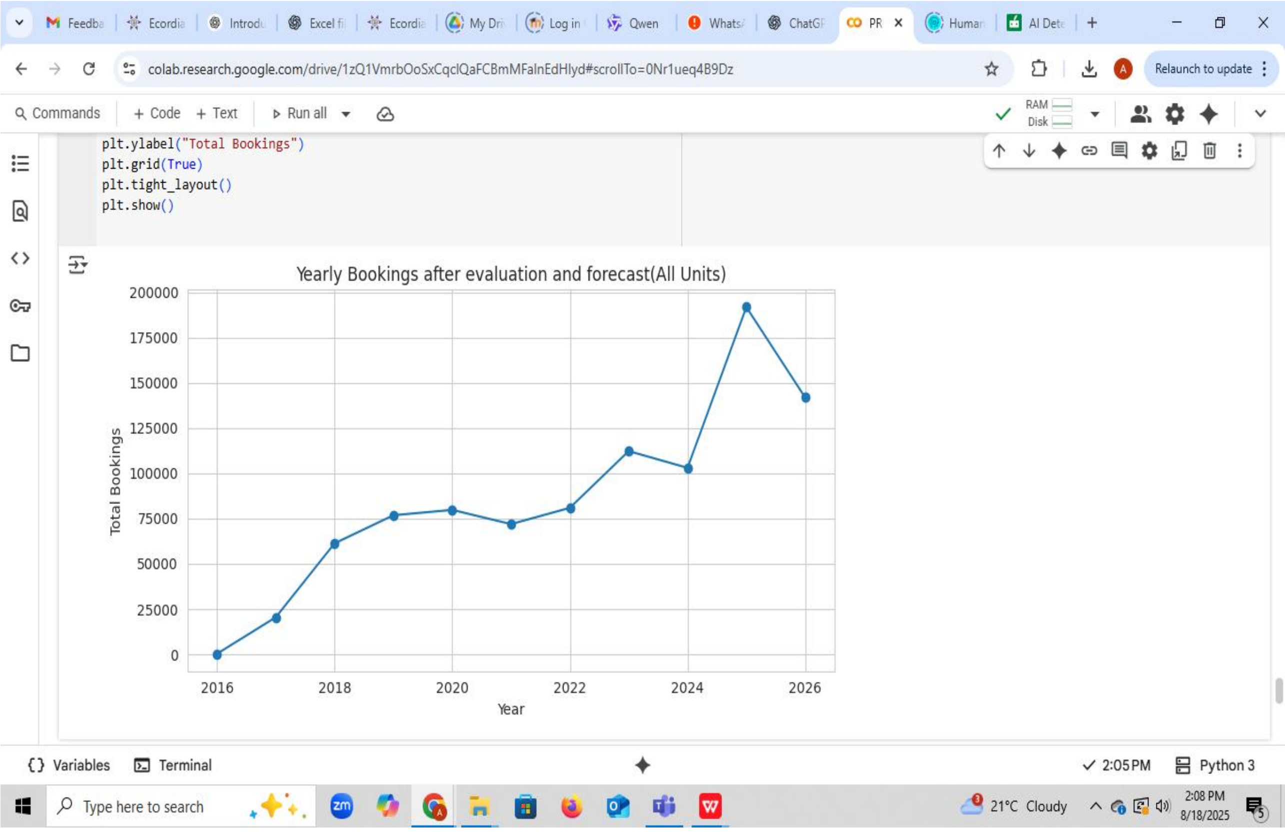


Fig. 49.3: Yearly booking plot with 2026 forecast inclusive (3)

In the process of forecasting for 2026 bookings the 2025 values (January to May) were scaled up to yearly data using a scaling factor ((12/5) x value). The line plot shows a spike in 2025 scaled. This can be accounted for as a result of some bookable units that started operation in 2025 with remarkable booking count and some that improved massively over previous years. Noting that the forecast for 2026 take into consideration historical bookings from previous years thereby curtailing its estimate unlike the estimates for 2025 that solely depend on its own values. This explains the decline in the trajectory from 2025 to 2026. Another point to note in the 2025 scaled data is that it does not fully take into consideration limiting factors like seasonality, availability and assessibility, technology factors among others.

4.2 Results

Summarisation of Research Question 1

What are the existing yearly booking patterns for the bookable units in the NHS buildings under review?

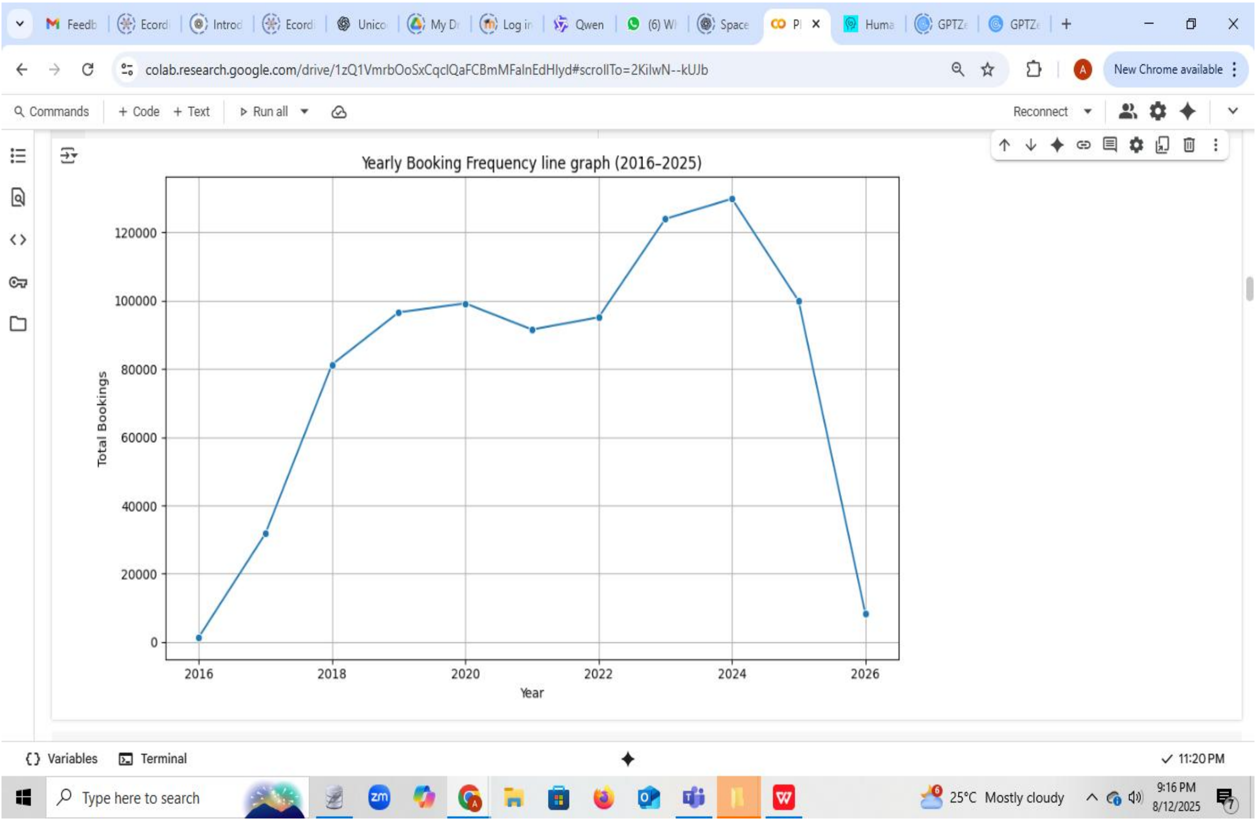


Fig. 17.2: Yearly booking line plot

From the yearly booking line plot above in fig. 17.2, the pattern follows a positive trajectory from 2016 until the covid period in 2020 where a decline in the trajectory was observed down to 2022 after the covid period. From 2022 the trajectory resumed its positive trend up to 2024 before another decline was observed in 2025. The decline observed in 2025 does not truly reflect the state of the pattern as the data captured for that year only cover from January to May of 2025 and data captured for 2026 are future bookings which does not really reflect the true data as bookings are still ongoing.

Summarisation of Research Question 2

At which (a) days of the week and (b) hours of the day does booking activity reach its peak?

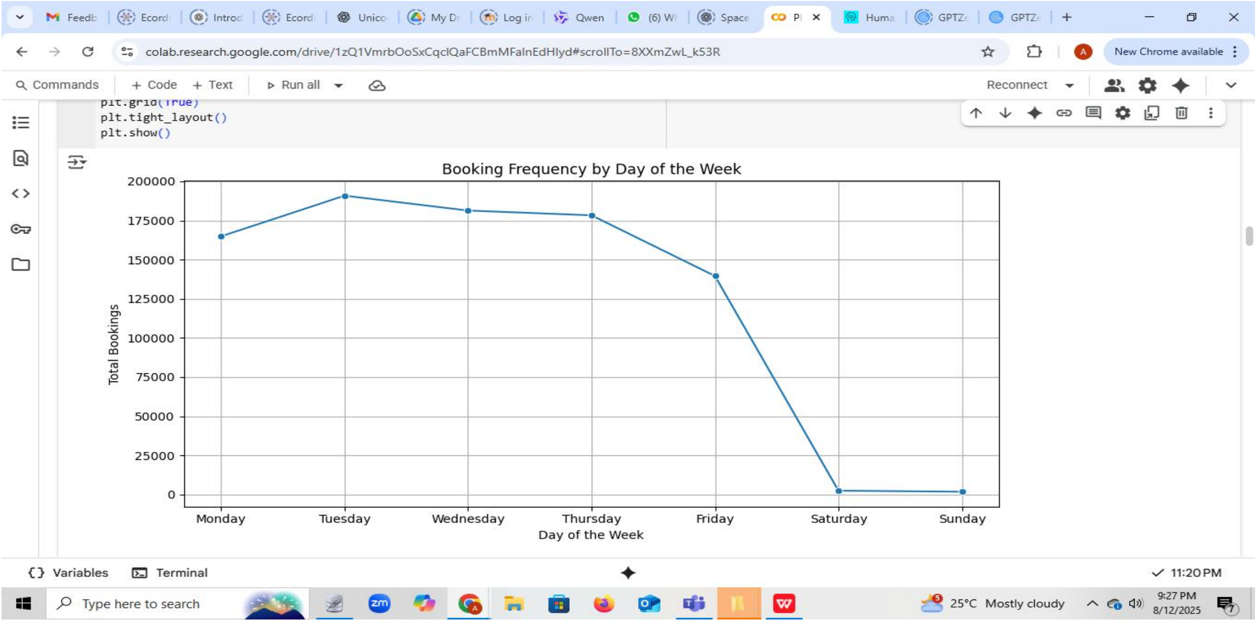


Fig. 18.2: Booking frequency by days of the week

a) From the booking frequency by days of the week plot, it can be observed that booking activity reach its peak on Tuesdays followed by Wednesdays with the least booking activity observed on Saturdays and Sundays

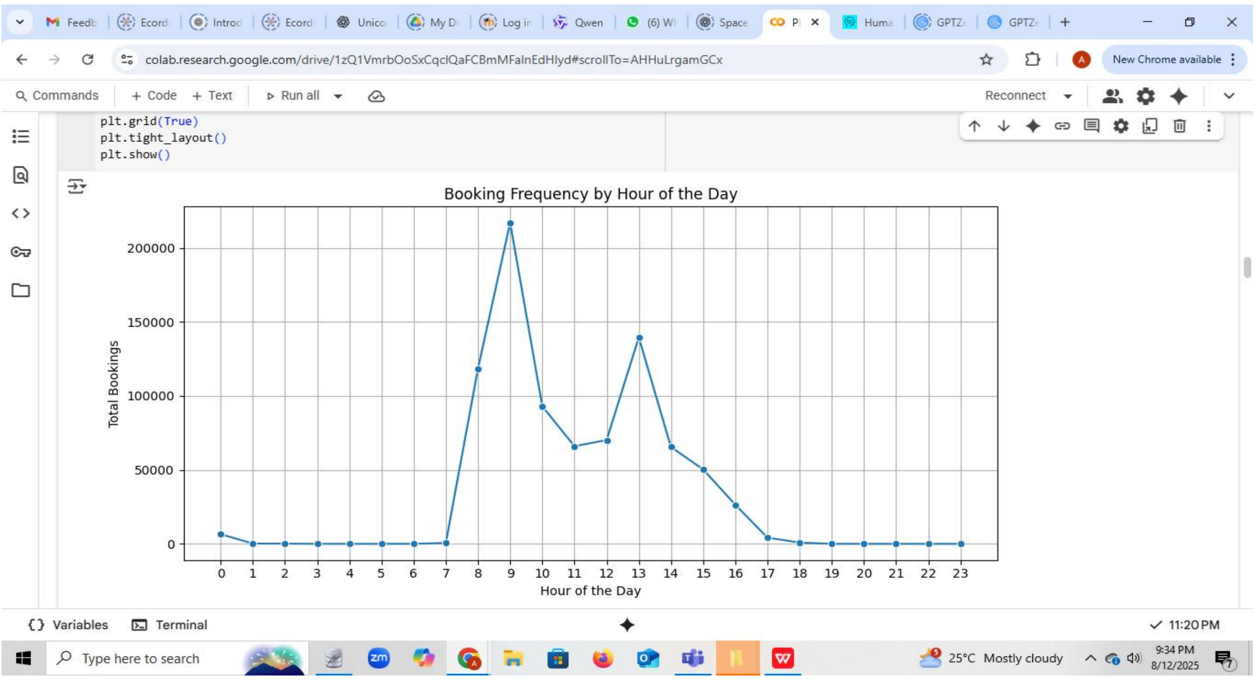


Fig. 19.2: Frequency distribution of bookings by hours of the day

b) From the frequency distribution of booking by hours of the day plot, it can be observed that booking activity reach its peak around 9:00 followed by 13:00 with the least booking activity observed from 18:00 down to 7:00 which is the time most of the buildings are closed.

Summarisation of Research Question 3

Based on the booking occasion history, which usage duration recorded the highest number of bookings?

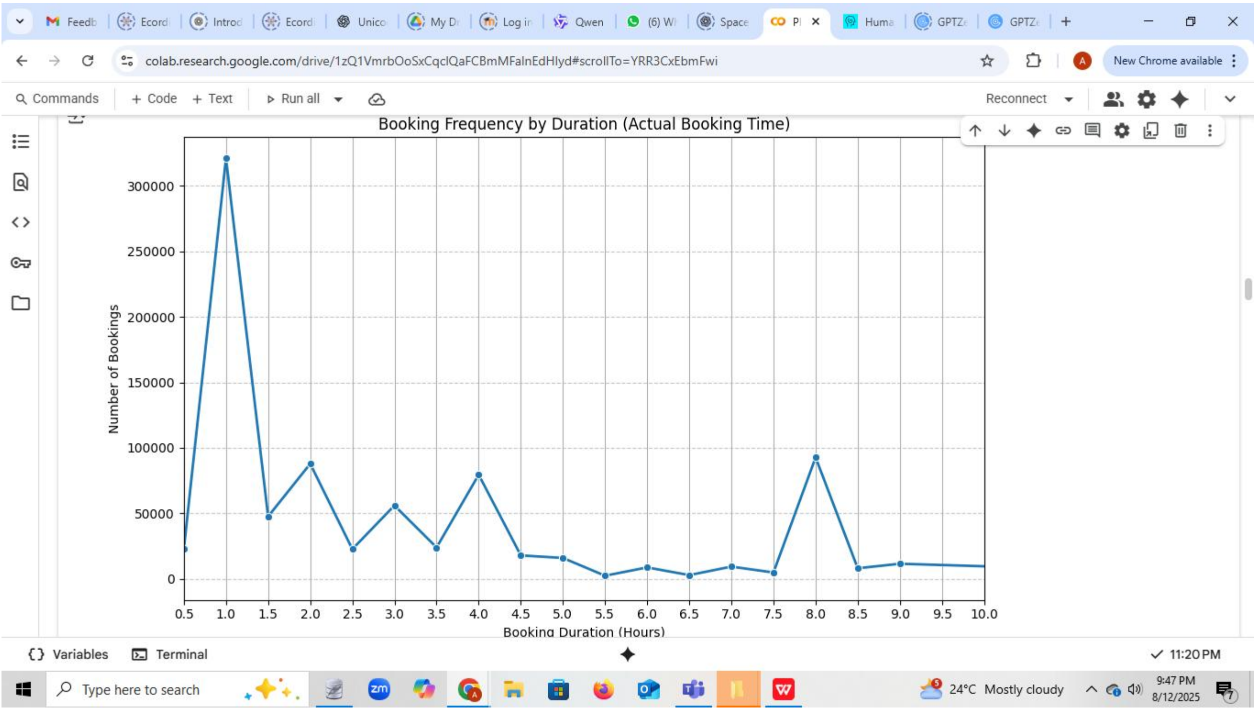


Fig. 20.2: Booking frequency by actual booking time used

The above plot shows that most bookings usage time is about one hour which most likely account for short meetings with the next most bookings usage time being about eight hours which most likely account for training sessions.

Summarisation of Research Question 4

What is the correlation between total yearly booking and utilisation rate of the bookable units?

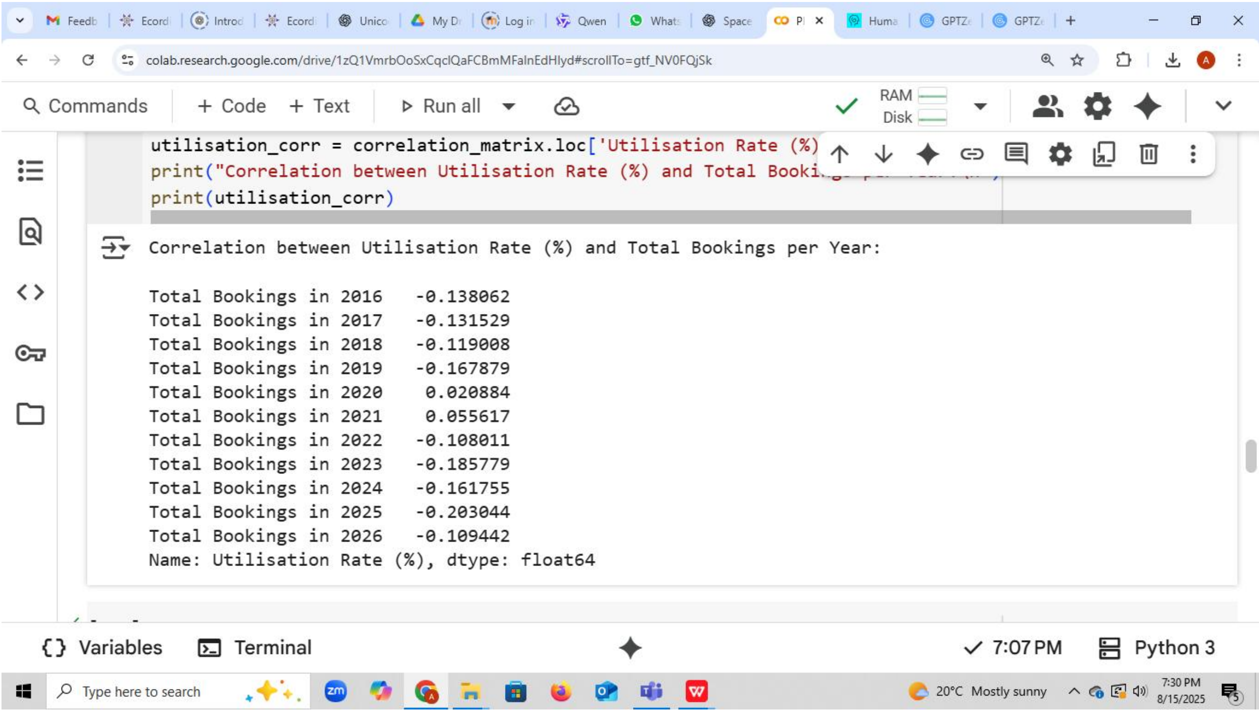


Fig. 29.2: Correlation summary

Fig. 30.2: Correlation Heatmap

The correlation analysis output from the correlation summary and heatmap plot shows that the utilisation rate has weak negative correlation with nine (9) yearly total booking and very week positive correlation with only two (2) yearly total booking with the heatmap additionally capturing all the correlations among all the yearly total bookings. This shows that satisfying the usage threshold from historical data does not guarantee that a bookable unit or building is still fit for the booking system operation as it may have a good utilisation rate yet not actively booked up to date.

Summarisation of Research Question 5

How many bookable units, buildings, and districts can be considered redundant within the years under review up to 2025, based on booking occasion data?

Category	Total Count
Districts	1
Buildings	7
Bookable Units	123

From the analysis carried out, after evaluation, it was observed that one district, seven buildings and one hundred and twenty three bookable units can be considered redundant within the years under review up to 2025. The reasons for categorising each of them as redundant are contained in the redundant tables of this study.

Summarisation of Research Question 6

What is the minimum number of buildings and bookable units currently required to sustain NHS booking system operations in the districts under review without negatively impacting service delivery?

District_id	Building_id	Bookable

		unit count
4	5	35
	6	20
	19	39
	22	2
	24	10
	46	6
5	7	17
	8	9
	25	13
6	9	11
	10	33
	11	14
	12	50
	14	3
	36	2
7	48	16
Total count **4**	**16**	**280**

Evaluated maywoods data summary

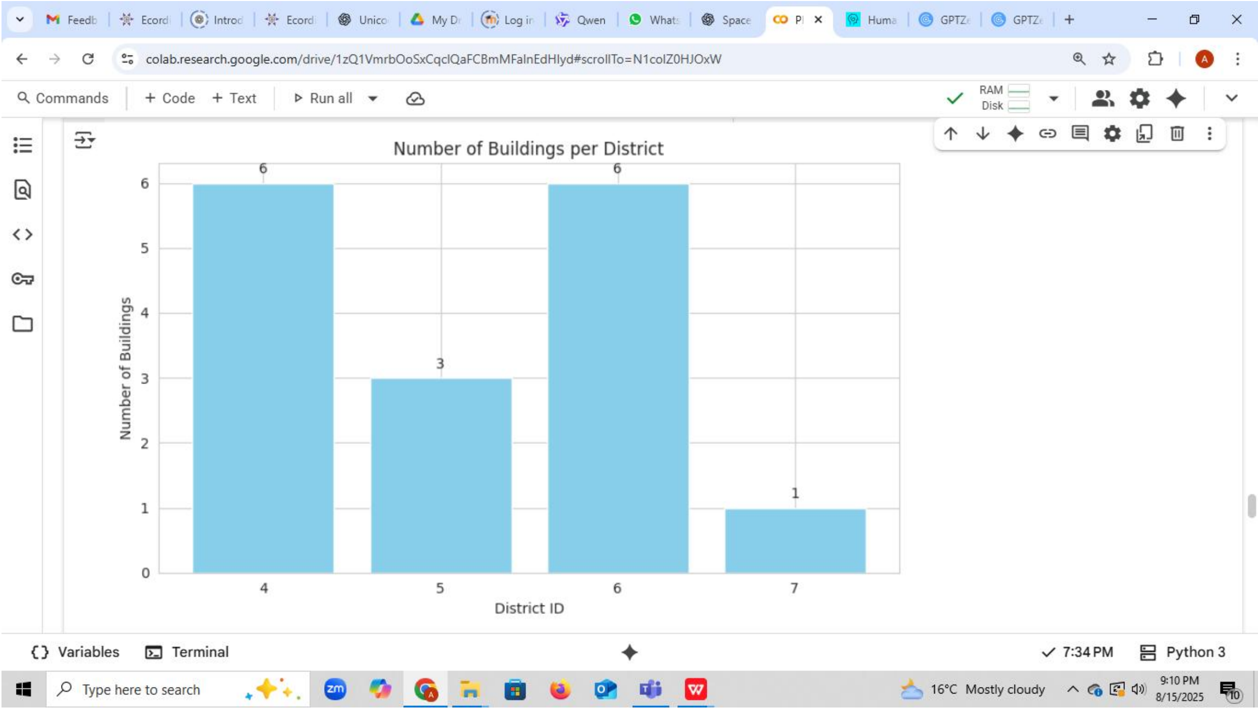


Fig. 34.2: District bar chart after evaluation

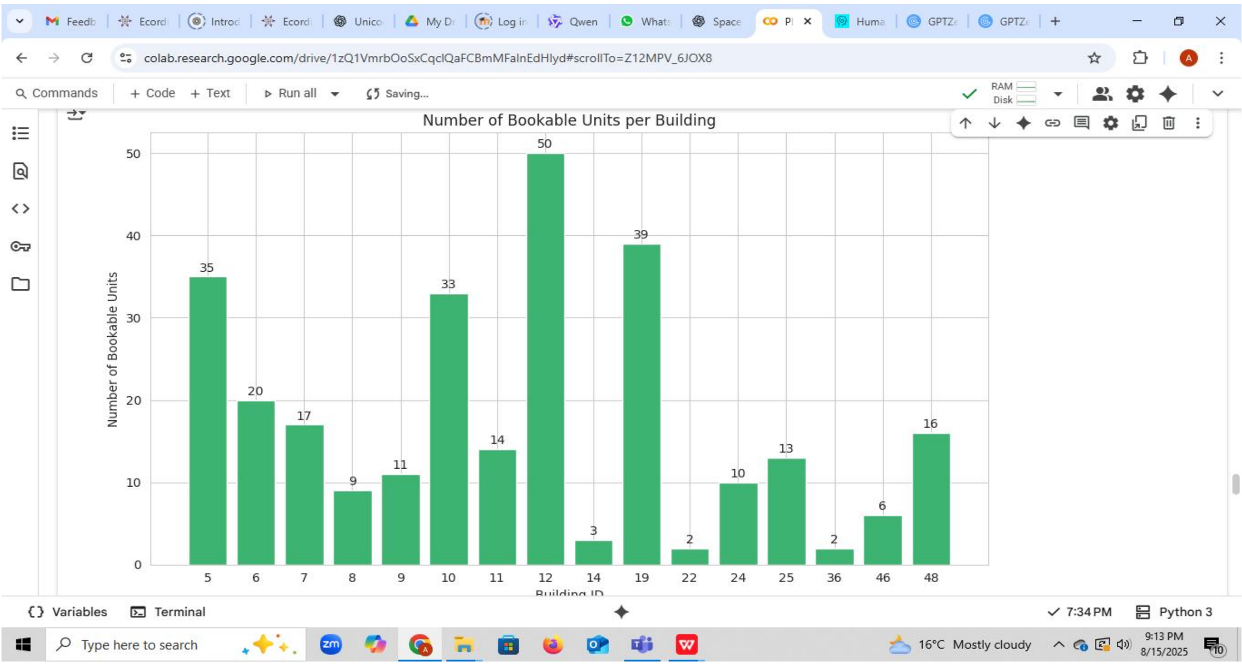


Fig. 35.2: Building bar chart after evaluation

The above table and charts shows the data summary after evaluation. It presents the distribution of the number of buildings and bookable units for each district that are required to sustain the NHS booking system operation without impacting on service delivery.

Summarisation of Research Question 7

(a) Which predictive model provides the best fit for forecasting the total yearly bookings of bookable units for 2026? ?

(b) What is the graphical comparison of the predicted values of best fit model with the actual values?

(c) What is the forecasted total yearly booking for bookable units in 2026?

MODEL	MEAN ABSOLUTE ERROR
Bayesian	128.83
Linear Regression	127.12
Random Forest	80.86
Gradient Boosting	89.45
Linear SVR	129.02

Model evaluation summary table

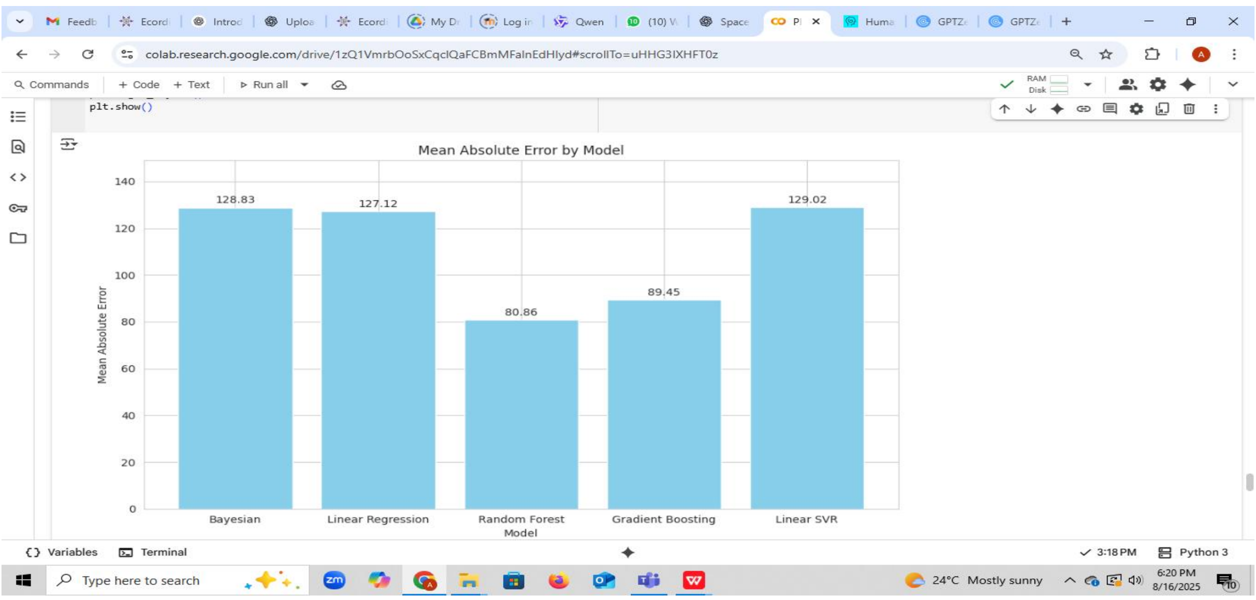


Fig. 46.2: Model evaluation chart

a) From the model evaluation summary table and chart, Random forest model provides the best fit for forecasting the total yearly booking for 2026 because it has the least mean absolute error compare to other models.

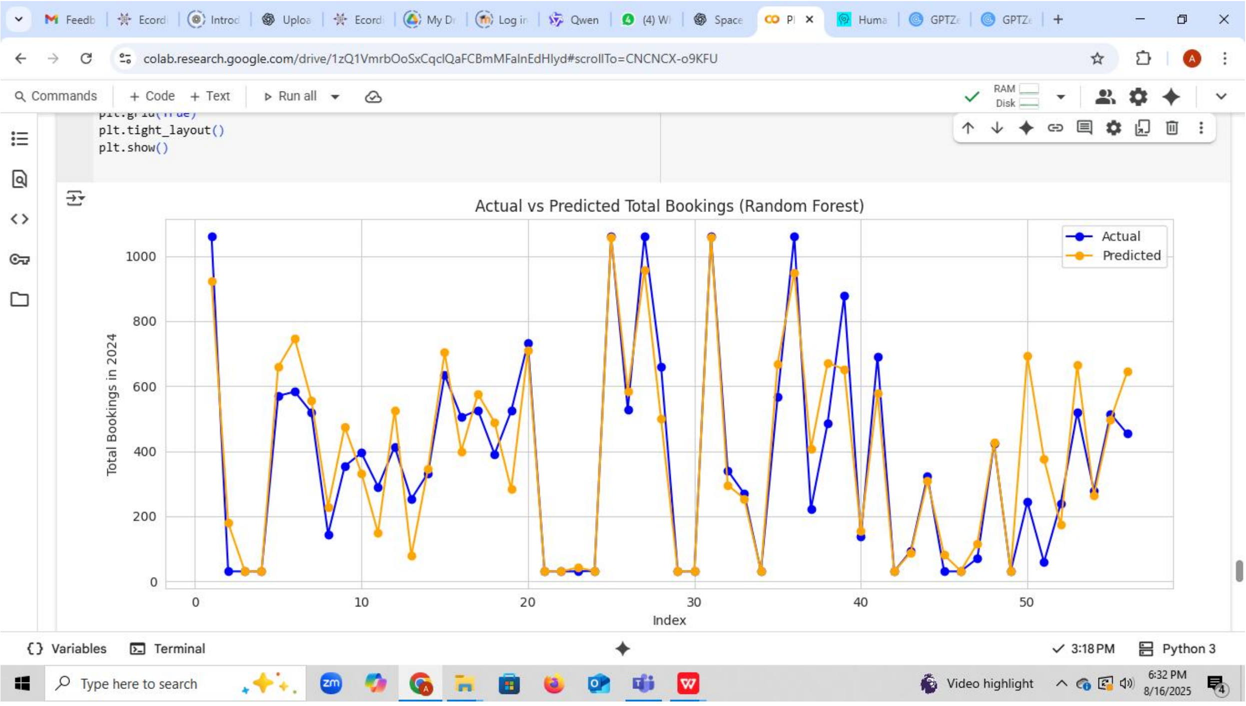
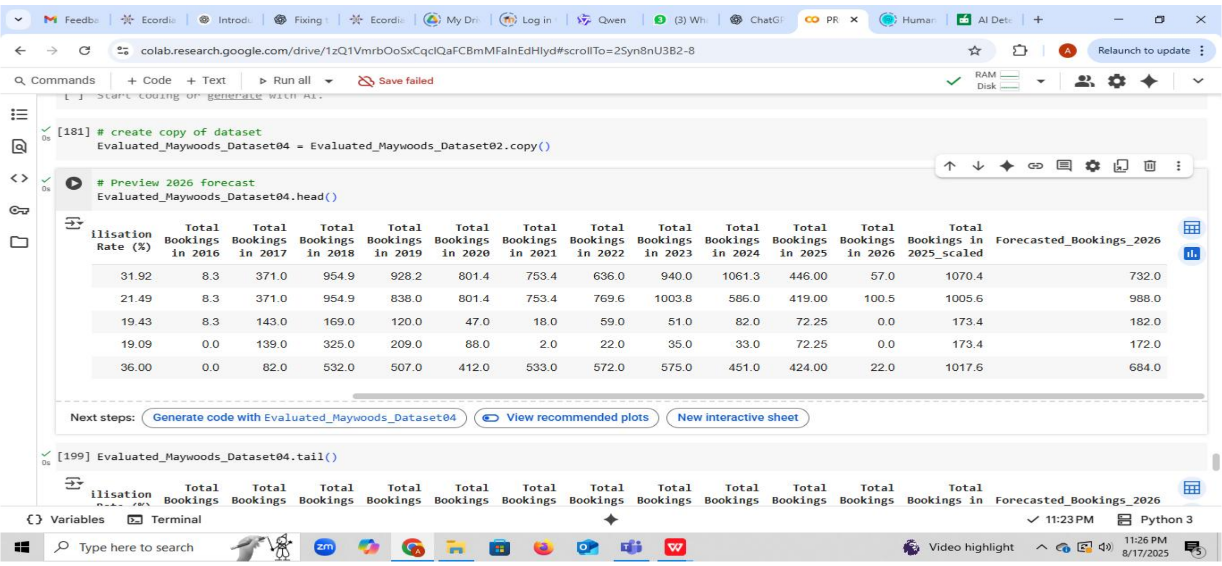


Fig. 47.2: Random forest model evaluation plot

b) *The Random forest model evaluation plot displays the comparison of the line patterns of the predicted and actual values. From the visualisation, the two lines have close pathways with few points showing significant difference.*



```
[181] # create copy of dataset
Evaluated_Maywoods_Dataset04 = Evaluated_Maywoods_Dataset02.copy()

# Preview 2026 forecast
Evaluated_Maywoods_Dataset04.head()
```

ilisation Rate (%)	Total Bookings in 2016	Total Bookings in 2017	Total Bookings in 2018	Total Bookings in 2019	Total Bookings in 2020	Total Bookings in 2021	Total Bookings in 2022	Total Bookings in 2023	Total Bookings in 2024	Total Bookings in 2025	Total Bookings in 2026	Total Bookings in 2025_scaled	Forecasted_Bookings_2026
31.92	8.3	371.0	954.9	928.2	801.4	753.4	636.0	940.0	1061.3	446.00	57.0	1070.4	732.0
21.49	8.3	371.0	954.9	838.0	801.4	753.4	769.6	1003.8	586.0	419.00	100.5	1005.6	988.0
19.43	8.3	143.0	169.0	120.0	47.0	18.0	59.0	51.0	82.0	72.25	0.0	173.4	182.0
19.09	0.0	139.0	325.0	209.0	88.0	2.0	22.0	35.0	33.0	72.25	0.0	173.4	172.0
36.00	0.0	82.0	532.0	507.0	412.0	533.0	572.0	575.0	451.0	424.00	22.0	1017.6	684.0

```
[199] Evaluated_Maywoods_Dataset04.tail()
```

Fig. 48.3: Prediction of total booking for 2026

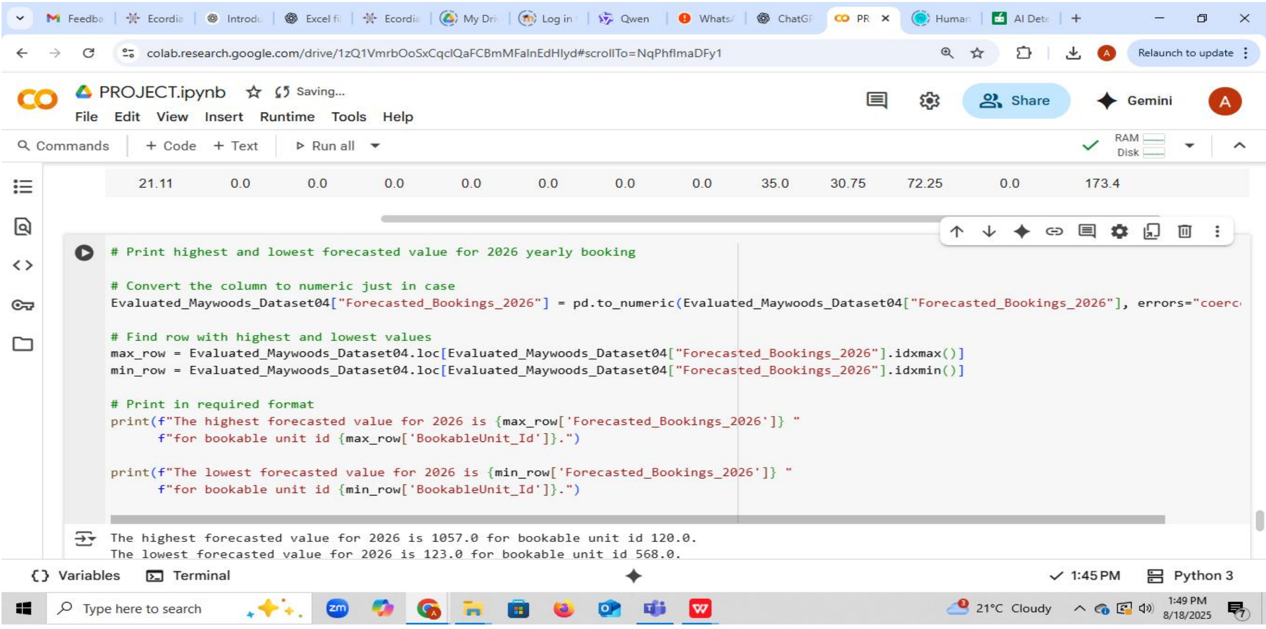


Fig. 50: Highest and Lowest forecast for 2026

c) From the Evaluated maywoods dataset containing the forecasted bookings for 2026, it presents bookable unit id 120. as the unit that will most likely have the highest booking of 1057 in 2026 and bookable unit id 658 as the bookable unit id that will most likely have the least booking of 123 in 2026.

4.3 Discussion

This study aimed to explore patterns of booking of NHS desk and meeting room space to identify redundancy, quantify sustainable capacity, and inform estate strategy through data-driven recommendations. The findings show unique patterns of booking over time together with significant inefficiencies in space utilisation, evidence towards a data-driven optimisation for organisations with limited resources like the NHS.

The annual trend in bookings was examined and shows a steady rising trend in demand from 2016 to the beginning of the COVID-19 pandemic in 2020, whereupon an instant decline occurred. This decline was sustained until 2022, after which bookings picked up gradually to 2024 before showing another apparent dip in 2025. The decline in this instance is because data were not yet complete and does not reflect a real decrease in demand. This suggests the need to read booking data in context and understand the effects of external shocks such as pandemics with the potential to reshape space use patterns.

At a more granular level, time-series analysis picks out Tuesdays and Wednesdays as the busiest days, with traffic increasing highest at 9:00 and also around 13:00. This captures typical work patterns where meetings tend to congregate at the start of the week and during middle office hours. Similarly, the usage duration analysis reveals that one-hour bookings as the most common, typical of brief meetings, followed by eight-hours booking which most likely account for training. These findings can be used by managers who wish to establish a balance between scheduling policy and room provision against usage patterns identified. The study also challenged the relationship between bookings and utilisation rates and concluded that high utilisation does not guarantee ongoing relevance in the bookings system. The weak and inconsistent correlations established reveal that reliance on use of utilisation levels is not sufficient to determine space efficiency. An example is that some units maintained historically high utilisation but were no longer actively booked within current years. This finding deserves greater dynamic and evidence-based measures of efficiency over reliance on traditional usage statistics.

Interestingly, the redundancy audit identified one (1) district, seven (7) buildings, and 123 bookable units that can be considered redundant without negatively impacting service delivery. By bundling capacity in 4 districts, 16 buildings, and 280 units, the NHS would be in a position to operate at an efficient and sustainable level, with properties still robust and reducing unnecessary overheads. The predictive modelling exercise that evaluated five algorithms demonstrated the potential for forecasting demand based on historical booking data. While Random Forest performed the best in accuracy, the modellings were designed as an aid tool and not as a final product, confirming the ability of data-driven methods to enhance estate decision making.

Overall, this research makes a contribution to demonstrating the application of SQL, python exploratory analysis, and machine learning in NHS book systems to provide actionable insights. It confirms there are significant opportunities for optimisation and that data-driven approaches can bridge the gap between theoretical models of space management and practical, operational gains within healthcare administrative environments.

4.4 Redundant buildings and bookable units summary

In the course of evaluating the booking data, it was discovered that one district, seven buildings and one hundred and twenty three bookable units have been redundant in the booking system up till 2025. Below are summary tables alongside their respective reasons for being classified as redundant;

REDUNDANT BUILDING_ID

District _id	Building _id	Building_Utilisation_Rate	Number of Bookable units	BookableUnit_id	BookableUnit_Utilisation_Rate	Reasons
6	35	9.38%	1	323	9.38%	Very low utilisation rate and actively booked in 2020, 2021 and 2022 only
5	20	14.52%	1	129	14.52%	Very low utilisation rate and actively booked in 2016 and 2017 only
4	39	18.20%	4	582, 592, 593, 594	13.89%, 12.04%, 33.33%, 13.73%	592 was booked 52 times and cancelled 51 times mostly by the trust in 2024. 593 was actively booked three times in a day for just one day in March 2024. 594 was booked only once in 2024. In 2024, 582 was booked 26 times and cancelled 9 times by

						user and only recorded one booking in January 2025. In summary, Building has only one booking in January 2025 with multiple cancelled bookings in 2024
4	47	25%	1	575	25%	It was booked only once in 2023 and thrice in early 2024
6	13	45.50%	10	294, 296, 297, 298, 299, 300, 301, 302, 303, 304	12.5, 43.25, 46.64, 42.31, 40, 51.45, 49.17, 80, 40, 49.84	Despite its remarkable usage rate compare to other buildings, it was only used from 2019 to 2020
5	17	58.80%	1	192	58.75	Despite its remarkable usage rate compare to other buildings, it was only used from 2018 to 2019

Table 6: Redundant building id table

REDUNDANT BOOKABLEUNIT_ID

District _id	Building _id	Total Number of Bookable units	Number of Bookable unit to be dropped	Bookable Unit_id	BookableUnit_ Utilisation_Rate (%)	Reasons
5	8	19	10	38	10	It was active in 2016 only
				39	5	it was booked only once in 2016
				40	14.98	It was active in 2016, 2017, 2018, 2019, 2022 and 2023 only
				41	18.64	It was booked only in 2016, 2017, 2018, 2019, 2022 and 2023
				42	82.03	Despite its high utilisation rate, it was effectively used only in 2016, 2017 and 2018
				44	7.5	It was booked only once in 2016 and 2019
				45	5	It was booked only once in 2016
				46	12.5	It was booked only once in 2016
				47	5	It was booked only in 2016
				48	7.5	It was booked only once in 2017
6	36	3	1	356	0	All bookings in 2025 were cancelled by user. Bookable unit recorded no usage

4	24	11	1	629	40.91	Recorded only two bookings in 2024 with no booking in 2025 despite being situated in a building with remarkable number of bookings.
4	22	3	1	194	23.02	It was only active from 2017 to 2021
6	9	12	1	125	50.35	It was only booked in 2016, 2017 and 2019
4	5	39	4	97	21.42	No booking recorded in 2024 and 2025
				101	9.19	Recorded booking only in 2016, 2017 and 2018
				111	19.89	It was booked only from 2016 to 2021
				157	17	It was booked only from 2016 to 2020
5	25	16	3	375	11.11	It was booked only once in 2022
				394	11.11	It was booked only once in 2023
				397	33.33	It was booked only once in 2022
6	10	41	8	134	10.42	It was actively booked only once in 2020
				166	2.78	It was actively booked only three times in 2017
				442	19.44	It was actively booked only three times in 2022

				445	8.33	It was actively booked only once in 2022
				459	25	It was actively booked only once in 2022
				461	2.08	It was actively booked only once in 2022
				495	25	It was actively booked only once in 2023
				508	14.58	It was actively booked only once in 2023
4	6	21	1	74	17.94	Last active booking was just twice in 2024. Any future booking to be redirected to any of the bookable units in the same building. Bookable unit can be repurpose for other uses.
6	12	55	5	212	18.5	It was actively booked only in 2018
				221	50.21	No booking recorded in 2025. Any future booking to be redirected to any of the bookable units in the same building.
				228	27.55	Last booking recorded was in 2021
				229	20.24	Last booking recorded was in 2021

<table>
<tr><td></td><td></td><td></td><td></td><td>514</td><td>23.2</td><td>Had no booking in 2025.</td></tr>
<tr><td rowspan="7">5</td><td rowspan="7">7</td><td rowspan="7">25</td><td rowspan="7">7</td><td>168</td><td>20</td><td>It recorded only one active booking in 2017</td></tr>
<tr><td>169</td><td>31.67</td><td>Last active booking of this unit was in 2023</td></tr>
<tr><td>170</td><td>23.33</td><td>Last active booking of this unit was in 2023</td></tr>
<tr><td>288</td><td>48.47</td><td>Last active booking of this unit was in 2023</td></tr>
<tr><td>310</td><td>41.67</td><td>Last active booking of this unit was in 2021</td></tr>
<tr><td>311</td><td>23.33</td><td>Last active booking of this unit was in 2021</td></tr>
<tr><td>312</td><td>16</td><td>Last active booking of this unit was in 2021</td></tr>
<tr><td rowspan="6">6</td><td rowspan="6">11</td><td rowspan="6">20</td><td rowspan="6">6</td><td>140</td><td>15</td><td>It was actively booked only in 2019</td></tr>
<tr><td>143</td><td>63.65</td><td>The last active booking of this unit was recorded in 2024</td></tr>
<tr><td>144</td><td>68.88</td><td>The last active booking of this unit was recorded in 2023</td></tr>
<tr><td>145</td><td>80</td><td>The last active booking of this unit was recorded in 2024</td></tr>
<tr><td>147</td><td>16.72</td><td>The last active booking of this unit was recorded in 2018</td></tr>
<tr><td>571</td><td>20.83</td><td>The last active booking of this</td></tr>
</table>

						unit was only two recorded in 2024
4	19	83	44	113	64.96	No booking recorded in 2025. Last booking was in 2022
				119	48.37	No booking recorded in 2025. Last booking was in 2024
				158	20.2	No booking recorded in 2025. Last booking was in 2023
				159	23.56	No booking recorded in 2025. Last booking was in 2024
				160	18.36	No booking recorded in 2025. Last booking was in 2022
				162	18.76	No booking recorded in 2025. Last booking was in 2023
				164	16.39	No booking recorded in 2025. Last booking was in 2022
				283	15.29	No booking recorded in 2025. Last booking was in 2024 and it was just one
				287	22.5	No booking recorded in 2025. Last booking was in 2024 and it was just one
				327	54.5	No booking recorded in 2025. Last booking was in 2022
				328	60.87	No booking recorded in 2025. Last booking was in 2022

				329	72.17	No booking recorded in 2025. Last booking was in 2022
				330	63.27	No booking recorded in 2025. Last booking was in 2022
				331	63.17	No booking recorded in 2025. Last booking was in 2022
				332	35.1	No booking recorded in 2025. Last booking was in 2021
				334	70.71	No booking recorded in 2025. Last booking was in 2021
				335	55.08	No booking recorded in 2025. Last booking was in 2022
				336	60.71	No booking recorded in 2025. Last booking was in 2022
				337	50	No booking recorded in 2025. Last booking was just one in 2021
				338	76.67	No booking recorded in 2025. Last booking was in 2021
				339	59	No booking recorded in 2025. Last booking was in 2021
				340	55	No booking recorded in 2025. Last booking was just two in 2020
				341	81.67	No booking recorded in 2025. Last booking was in 2020

				342	2.5	No booking recorded in 2025. Last booking was just one in 2020
				343	45	No booking recorded in 2025. Last booking was just one in 2022
				344	50.83	No booking recorded in 2025. Last booking was just one in 2022
				346	77.27	No booking recorded in 2025. Last booking was in 2022
				348	64.64	No booking recorded in 2025. Last booking was just one in 2023
				349	2.5	No booking recorded in 2025. Last booking was just two in 2021
				350	40	No booking recorded in 2025. Last booking was in 2022
				352	24.5	No booking recorded in 2025. Last booking was just one in 2024
				410	87.5	No booking recorded in 2025. Last booking was in 2022
				511	10	No booking recorded in 2025. Last booking was just one in 2022

				523	50	No booking recorded in 2025. Last booking was in 2022
				525	62.5	No booking recorded in 2025. Last booking was in 2022
				526	31.87	No booking recorded in 2025. Last booking was in 2022
				527	73.19	No booking recorded in 2025. Last booking was in 2022
				528	80	No booking recorded in 2025. Last booking was just one in 2022
				530	37.5	No booking recorded in 2025. Last booking was just two in 2022
				531	58.5	No booking recorded in 2025. Last booking was in 2022
				532	80	No booking recorded in 2025. Last booking was in 2022
				533	70	No booking recorded in 2025. Last booking was just one in 2022
				552	40.71	No booking recorded in 2025. Last booking was in 2024
				559	40	No booking recorded in 2025. Last booking was just one in 2022

Table 7: Redundant bookable unit table

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The project was tasked with examining and enhancing the NHS booking system by combining twenty three datasets, data cleaning and pre-processing, exploratory analysis, and predictive modeling. It was discovered that the system initially comprised five (5) districts, twenty three (23) buildings, and four hundred and three (403) bookable units, but a high percentage of these assets were discovered to be redundant. Upon evaluation, the NHS booking system proved to be feasible with four (4) districts, sixteen (16) buildings, and two hundred and eighty (280) bookable units, being the barest minimum setup for which continuous effective provision of service could be ensured. Principal booking patterns derived from exploratory work were trends in time (Tuesday and 9:00 hours being most in demand), predominance of eight-hour and one-hour booking intervals, and the weak association of utilisation rates and bookings figures. These findings underscore the importance of employing more than one approach when measuring resource efficiency instead of utilisation alone. Secondly, testing also confirmed that decommissioning one district, seven buildings, and one hundred and twenty three bookable units would not negatively impact service outcomes, thus advancing evidence-based justification for operational streamlining. From a predictive point of view, the study used five machine learning algorithms and established that Random Forest regressor gave the best precision with minimal mean absolute error. The model was then utilised to predict bookings in 2026 and generated a scenario range of 123 to 1,057 bookings per unit. The findings provide a forward looking aspect to the analysis and give methodological template for continued monitoring and planning. In all, the results name three of the project's most important contributions: (1) establishing the data-driven minimum capacity threshold standard of the NHS booking system, (2) facilitating actionable knowledge for redundancy reduction and demand management, and (3) confirming the worth of predictive analytics as a resource-planning instrument. It also specifies its limitations, i.e., unavailability of 2025 complete data and utilising modelled rather than observed data for 2026, and proposes the need for frequent model retraining, proper data governance, and continuous stakeholder involvement. Lastly, this project provides a robust but pragmatic solution to the development of the NHS booking system. By employing extensive data cleansing, exploratory data analysis, redundancy analysis, and predictive modeling, it maximises both day-to-day effectiveness and strategic vision. Such recommendations offered in this report, when implemented, will provide the development of a leaner, healthier, and less expensive booking system that remains responsive to staff and organisational requirements but also responsive to the issues that it is likely to encounter in the future.

In summary, analysis confirms that the NHS booking system is able to function suitably with a reduced footprint of 280 bookable units across four districts and sixteen buildings. Strategic retirement of underused assets, along with more complex utilisation schemes, enhanced forecasting techniques, and better demand management will provide an improved more efficient and less expensive booking system that does not compromise the delivery of service. Ongoing monitoring, periodic model retraining, stakeholder engagement, and policy book revision for ensuring resilience and flexibility should be the subject of future research

5.2 Results summary

Table 6 below shows the summary of the research questions alongside the corresponding answers

S/N	RESEARCH QUESTIONS	CORRESPONDING ANSWERS
1	What are the existing yearly booking patterns for the bookable units in the NHS buildings under review?	*From the yearly booking line plot above in fig. 17.2, the pattern follows a positive trajectory from 2016 until the covid period in 2020 where a decline in the trajectory was observed down to 2022 after the covid period. From 2022 the trajectory resumed its positive trend up to 2024 before another decline was observed in 2025. The decline observed in 2025 does not truly reflect the state of the pattern as the data captured for that year only cover from January to May of 2025 and data captured for 2026 are future bookings which does not really reflect the true data as bookings are still ongoing.*
2a	At which days of the week does booking activity reach its peak?	*a) From the booking frequency by days of the week plot, it can be observed that booking activity reach its peak on Tuesdays followed by Wednesdays with the least booking activity observed on Saturdays and Sundays*
2b	At which hours of the day does booking activity reach its peak?	*a) From the frequency distribution of booking by hours of the day plot, it can be observed that booking activity reach its peak around 9:00 followed by 13:00 with the least booking activity observed from 18:00 down to 7:00 which is the time most of the buildings are closed.*
3	Based on the booking occasion history, which usage duration recorded the highest number of bookings?	*From the actual booking hours plot, most bookings usage time is about one hour which most likely account for short meetings with the next most bookings usage time being about eight hours which most likely account for training sessions.*

4	What is the correlation between total yearly booking and utilisation rate of the bookable units?	*The correlation analysis output from the correlation summary and heatmap plot shows that the utilisation rate has weak negative correlation with nine (9) yearly total booking and very week positive correlation with only two (2) yearly total booking with the heatmap additionally capturing all the correlations among all the yearly total bookings. This shows that satisfying the usage threshold from historical data does not guarantee that a bookable unit or building is still fit for the booking system operation as it may have a good utilisation rate yet not actively booked up to date.*
5	How many bookable units, buildings, and districts can be considered redundant within the years under review up to 2025, based on booking occasion data?	*From the analysis carried out, after evaluation, it was observed that 1 district, 7 buildings and one 123 bookable units can be considered redundant within the years under review up to 2025. The reasons for categorising each of them as redundant are contained in the redundant tables of this study*
6	What is the minimum number of buildings and bookable units currently required to sustain NHS booking system operations in the districts under review without negatively impacting service delivery?	*From the analysis carried out, a minimum of 4 districts, 16 buildings and 280 bookable units are required to sustain the NHS booking system operation without impacting on service delivery.*
7	Which predictive model provides the best fit for forecasting the total yearly bookings of bookable units for 2026? ?	*From the model evaluation summary, Random forest model provides the best fit for forecasting the total yearly booking for 2026 because it has the least mean absolute error compare to other models*

Table 8: Research questions and answer summary

5.3 Limitations of the study

Incomplete data: Bookings for 2025 only through January to May, and 2026 is forecasted, which limits the validity of trends. Future forecasts must be for complete years.

Model assumptions: Forecast assumes historical demand patterns are stable. Shock events (pandemics, policy shocks) can make forecasts less reliable.

Redundancy classification: Some units, like Bookable Unit 42, have high historical utilisation but are currently inactive. Delisting these units will dampen niche or specialist demand.

Potential disruption: Reducing capacity will restrict operational flexibility in the face of unforeseen peaks in demand. Maintaining a buffer of redundant units on standby mitigates this.

Data quality: Preprocessing and merging errors done while merging datasets can influence conclusions. SQL merging and cleaning needs to be well documented and reproducible.

5.4 Recommendations

The NHS booking system analysis, following the integration and analysis of twenty-three datasets, has provided a clear picture of the current usage, redundancy, and predictive forecasts. The project established that following data cleansing, pre-processing and evaluation, the NHS booking system infrastructure could ideally operate at four (4) districts, sixteen (16) buildings, and 280 bookable units, the lowest possible capacity without impacting service delivery. This decrease follows identification of a single redundant district, seven redundant buildings, and 123 redundant bookable units.

Key Recommendations

1. Retire redundant buildings and bookable units

The redundancy figures show one district, seven buildings, and 123 units redundant. My recommendation is to retire these and plan around 280 bookable units available.

Overheads will be lowered and efficiency maximised without impacting delivery of service. Some of these can be retained in "standby" to allow for quick reactivation in the event of emergencies or increased demand.

2. Examine the inactive district

One of the districts has been dormant since 2016. Further exploration is needed to understand the underlying reasons such as for demographic change, policy change, or operational inefficiencies. This would guide whether the district is discarded permanently, recycled, or repurposed for other health service delivery.

3. Reform utilisation measures

The low (predominantly negative) correlation between utilisation rates and number of bookings per year demonstrates that utilisation rate by itself is not an adequate indicator of space demand. A composite utilisation index that combines booking frequency, booking length, and utilisation percentage is needed. The broader measure will give a more realistic representation of true demand and prevent units from being prematurely labelled as obsolete.

4. Enhance forecasting practice

Random Forest model was found to generate the most precise forecasts, with 2026 forecasts ranging from 123 to 1,057 bookings per bookable unit on average. The model must be retrained annually using new data in order to ensure reliability of forecasts, it is recommended. Scenario planning forecasting (best, worst, and most likely case) must be implemented in order to allow NHS managers to plan under uncertainty. A dashboard to display forecast vs actual bookings would also provide transparency in operations.

5. Safeguard delivery of services

If there is downsizing, while the system will cope with 280 units, a cautious policy of reduction must be adhered to. Peak demand stress-testing simulations must be conducted to make certain that reduced capacity does not result in services facing bottlenecks. Regular consultation with NHS staff and district managers must ensure frontline service delivery is uninterrupted by the merger.

6. Strengthen temporal allocation strategies

Since Tuesday and 9:00 are peaks in booking demand, off-peak demand redistribution can be attempted by NHS managers by off-peak booking incentives. Flexible booking slots, or automated reminders can be introduced to end queuing and redistribute usage more evenly over the week.

7. Optimise booking duration policies

One-hour slots are the most frequent, and eight-hour slots are next. Regulations must be adjusted so that short time slots can be reserved for high-demand equipment and long blocks can be managed so that underutilisation does not occur. A cap or limit on long-duration bookings would optimise efficiency.

8. Stakeholder engagement and training

Front line staff, district managers, and administrators need to be engaged in redundant classification checking and learning from booking trends. Forecast interpretation training and utilisation metrics will make decision makers consistent and effective in using the system data.

9. Develop a monitoring dashboard

Establish a real time monitoring dashboard measuring bookings, utilisation, peaks, and forecast accuracy. This will introduce transparency, allow for data driven decisions, and provide early intervention in case of a potential risk to service delivery.

10. Future proof the booking system

To guard against more long-term shifts in demand, NHS needs to establish a cycle of ongoing improvement: periodic data audits, regular retraining of predictive models, and integration of exogenous information (e.g., population demographics, policy initiatives). This tests flexibility and robustness of the booking system.

5.5 Future Work

Even more crucial than this project's findings and recommendations are the numerous paths for future work to extend and improve. First, future studies must incorporate complete booking statistics for 2025 as well as actual booking statistics for 2026 when published, for the sake of confirming and refining current projections. This would increase the legitimacy of temporal trend analysis and provide a more accurate baseline on which forecasting models may be constructed.

Secondly, scope exists for enhancing the predictive modelling approach. While Random Forest regressor performed best in this research, there can be experiments with other more advanced techniques such as deep learning models, ensemble hybrid models, or time-series specific models such as LSTMs or Prophet in future studies. It may also enhance predictive power and robustness with comparative study of these models.

Thirdly, the project has also brought to light the limitations of relying on utilisation rates as the only means of measuring efficiency. Further research needs to formalise improving the development of a composite utilisation index that involves booking frequency, length of stay, and occupancy. Pilot testing the index across a number of years and in various contexts might legitimize its value and make it an accepted standardised performance measure in healthcare resource management.

Fourth, stakeholder engagement needs to be explored more rigorously. Creating an understanding of NHS managers', frontline practitioners', and patients' perceptions about booking redundancies and utilisation could provide us with more robust, context-dependent knowledge and more strongly link analytical outputs to working conditions.

Finally, broadening the analytic base to include exogenous factors such as demographic change, population growth, or policy changes in health care would provide a more accurate depiction of demand. Addition of these variables into forecasting models has the potential to allow for scenario planning based on long-range strategic concerns.

Together, such research avenues would not only enhance the current analysis but extend its scope, turning the NHS booking system into data-driven and strong according to shifting healthcare requirements.

5.6 Personal Reflection

This project has been challenging and worth it. I had underestimated how complex it was to consolidate twenty-three datasets into a single strong source of truth. Data preprocessing and cleaning took more time than I anticipated, but the test of patience, diligence, and transparency value in completely explaining every step was invaluable. It also made me value more the role that data quality can have on the validity of findings. Exploratory analysis proved particularly revealing. Discovering that 9:00 a.m. and Tuesdays were busiest to book, and one-hour bookings most common, allowed me to identify data back to activity by day of week in NHS practice. I learned to appreciate the reality that not all numbers tell the whole story, for example utilisation rates were useful at first but confusing when taken in a vacuum. That helped to inform my experience that quality analysis often entails stripping assumptions away and examining things from a whole host of vantage points. The most technical was construction and development of predictive models and was the most exhilarating part of the project. I became accustomed to writing with machine learning tools and how to cross-model in an impartial manner.. A balance of trust in data and cautious scepticism has been one of my greatest personal learning. Aside from the technical need, I benefited a lot from project management and persistence. There were frustrations of code not performing as I intended or results being the opposite of what my assumptions indicated. But having solutions to those just made me more resolute and skilled at solving problems. More importantly, I realized that university research is not just producing results but thinking critically about procedures, boundaries, and implications. This project has transformed me into a better technologist in SQL, data analysis, and machine learning. I am now more confident about navigating through tough hurdles, presenting results in plain English, and making quality recommendations. I leave this project with a deeper sense of the issue of achieving healthcare resource balancing and with a demand for continued learning and development as both practitioner and researcher. Through this project, I was inspired to conclude with the phrase: Data is a binocular that helps us see beyond perceptions and assumptions.

REFERENCES

Aw, S.F. and Rohayu, O., 2020. How workplace condition affects healthcare workers: Assessment at tertiary hospitals. *Malaysian Journal of Public Health Medicine*, *20*(Special1), pp.251-254.

Bodin Danielsson, C. and Theorell, T., 2019. Office employees' perception of workspace contribution: A gender and office design perspective. *Environment and Behavior*, *51*(9-10), pp.995-1026.

Cawood, T., Saunders, E., Drennan, C., Cross, N., Nicholl, D., Kenny, A., Meates, D. and Laing, R., 2016. Creating the optimal workspace for hospital staff using human centred design. *Internal Medicine Journal*, *46*(7), pp.840-845.

Fleming, S., Apps, N., Harbon, P. and Baldock, C., 2012. Assessing space utilisation relative to key performance indicators–how well, not how much, space is used. *Journal of Higher Education Policy and Management*, *34*(5), pp.503-515.

Ghasemkhani, A., Torabi, S.A. and Hamid, M., 2023. A hybrid simulation and optimisation approach for capacity allocation of operating rooms under uncertainty. *International Journal of Systems Science: Operations & Logistics*, *10*(1), p.2244423.

Harris, J., 2019. Workplace design+ space utilisation= talent retention: How understanding space usage can fund workplace transformations to deliver the offices our talent really wants. *Corporate Real Estate Journal*, *8*(4), pp.312-338.

Harris, R., 2015. The changing nature of the workplace and the future of office space. *Journal of Property Investment & Finance*, *33*(5), pp.424-435.

Kim, J., Candido, C., Thomas, L. and De Dear, R., 2016. Desk ownership in the workplace: The effect of non-territorial working on employee workplace satisfaction, perceived productivity and health. *Building and Environment*, *103*, pp.203-214.

Lange, S., Häne, E. and Windlinger, L., 2022. Workplace utilization in hospitals: a study of space efficiency potentials. *The Evolving Scholar*.

Nejati, A., Shepley, M., Rodiek, S., Lee, C. and Varni, J., 2016. Restorative design features for hospital staff break areas: A multi-method study. *HERD: Health Environments Research & Design Journal*, *9*(2), pp.16-35.

Peavey, E. and Cai, H., 2020. A systems framework for understanding the environment's relation to clinical teamwork: A systematic literature review of empirical studies. *Environment and Behavior*, *52*(7), pp.726-760.

Rapport, F., Auton, E., Cartmill, J., Braithwaite, J., Shih, P., Hogden, A. and Clay-Williams, R., 2019. Fit for purpose? OrganisationaL prOdUctivity and woRkforce wellbeIng in workSpaces in Hospital (FLOURISH): a multimethod qualitative study protocol. *BMJ open*, *9*(4), p.e027636.

Sadatsafavi, H., Walewski, J. and Shepley, M.M., 2015. Factors influencing evaluation of patient areas, work spaces, and staff areas by healthcare professionals. *Indoor and Built Environment*, *24*(4), pp.439-456.

Shetty, R.S., Kamath, G.B., Rodrigues, L.L., Nandineni, R.D. and Shetty, S.R., 2024. The Impact of the Physical Environment on Staff in Healthcare Facilities: A Systematic Literature Review. *Buildings*, *14*(9), p.2773.

Sutton, L., Tarrant, C., Willars, J., Coats, T., Simmonds, M., Mclean, D., Boyle, A., Dreesbeimdiek, K., Richter, S., Oyedijo, A. and Roland, D., 2025. How do staff work in NHS hospital operations management meetings to support resilience in everyday service delivery? A qualitative study. *BMC Health Services Research*, *25*(1), p.113.

Tabak, V., 2009. User simulation of space utilisation: system for office building usage simulation.

Tagliaro, C., Zhou, Y. and Hua, Y., 2021. A change in granularity: measure space utilization through smart technologies. *Facilities*, *39*(1/2), pp.64-79

Voordt, T.V.D. and Jensen, P.A., 2023. The impact of healthy workplaces on employee satisfaction, productivity and costs. *Journal of Corporate Real Estate*, *25*(1), pp.29-49.

Water, T., Wrapson, J., Reay, S. and Ford, K., 2018. Making space work: Staff socio-spatial practices in a paediatric outpatient department. *Health & Place*, *50*, pp.146-153.

Workero, 2025. *Meeting room image from Workero meeting room booking system*. Workero [online]. Available at: https://www.workero.com/meeting-room-booking-system/ [Accessed 26 August 2025].

BIBLIOGRAPHY

Abbou, B., 2022. Optimizing Operating Room Utilisation—A Predictive Model. *Data*, 6(3), p.76. DOI:10.3390/data6030076.

Baxendale, J. (2024) Optimising the use of healthcare estate. Baxendale Blog, 11 June. Available at: https://blog.baxendale.co.uk/optimising-the-use-of-healthcare-estate (Accessed: 9 July 2025).

Clare Young (2024) Optimising estate management: a key to NHS cost savings. Clarity Consulting Associates Ltd, 8 June. Available at: https://www.ccal.co.uk/post/optimising-estate-management-a-key-to-nhs-cost-savings (Accessed: 11 July 2025).

Einola, K.A., 2025. Optimizing office space: the role of passive occupancy. *Journal of Corporate Real Estate*. DOI:10.1108/JCRE-07-2024-0023.

Mosteiro-Romero, M., Miller, C., Chong, A. and Stouffs, R., 2022. Elastic buildings: calibrated district-scale simulation of occupant-flexible campus operation for hybrid work optimization. *arXiv preprint*. arXiv:2210.06124.

Norouziasas, A., 2024. Impact of space utilization and work time flexibility on energy use and thermal comfort. *Energy and Buildings*, 310, 112420. DOI:10.1016/j.enbuild.2024.112420.

Orhan, F. & Kurutkan, M.N., 2025. Predicting total healthcare demand using machine learning: separate and combined analysis of predisposing, enabling, and need factors. *BMC Health Services Research*, 25, Article 366. DOI:10.1186/s12913-025-12502-5.

RESEARCH SCHEDULE

This 12-week research plan was formulated at the commencement of this study and was used as a guide for planning and implementation of the project on **Evaluation of Room and Desk Booking Trends for Strategic Recommendations in NHS Facilities.**

The schedule outlines the main research stages, goals, and expected timelines.

Research Stages	Goals	Schedule
Exploratory Study for Thesis Topic / Research Proposal Preparation	Review scholarly literature on space utilisation, booking systems, and data-driven optimisation. Identify research gaps and establish the rationale for the study. Draft proposal with aims, objectives, and contribution.	Week One
Proposal Submission	Submission of research proposal	Week Two
Literature Review	Examine further scholarly works on methodologies, techniques, and data-driven approaches to space management. Identify deficiencies in the literature to strengthen the study's contribution.	Week Three – Four
Methodology	Collect and merge NHS booking datasets using SQL. Define data preprocessing steps, feature engineering, and exploratory analysis strategy in Python. Document predictive modelling framework.	Week Five – Six

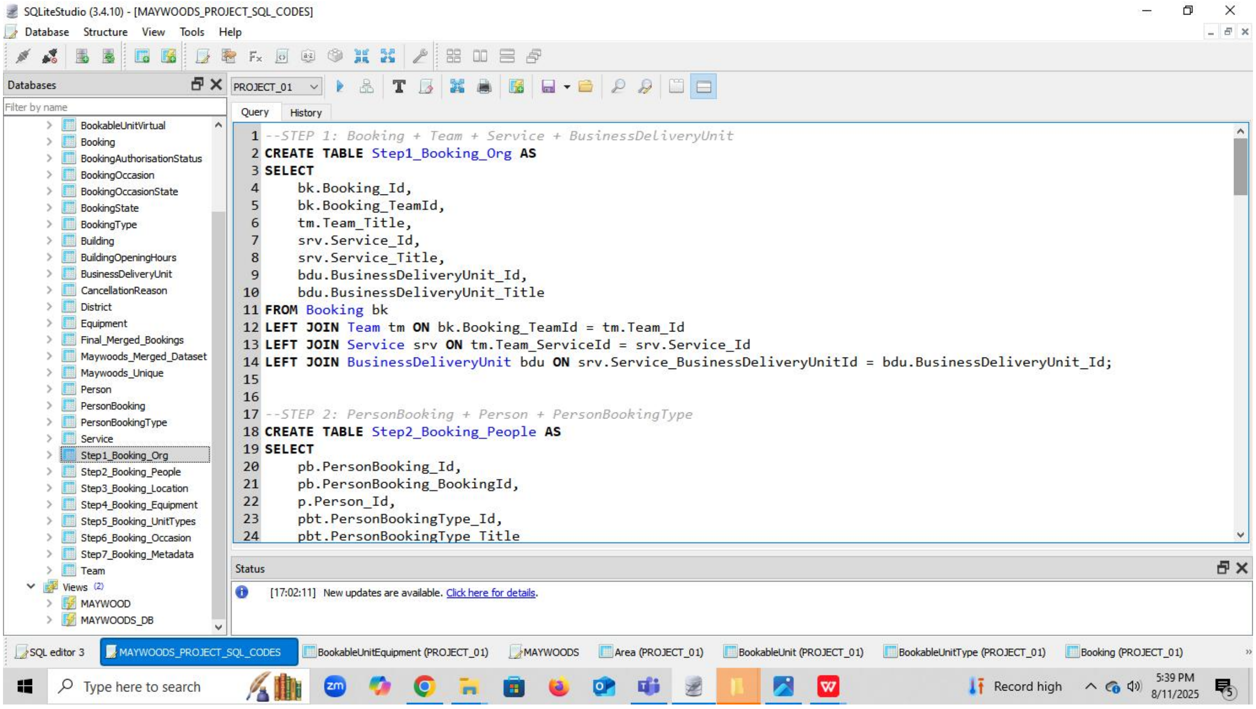
Implementation	Apply preprocessing and cleaning to datasets. Conduct exploratory data analysis (EDA) to detect booking patterns, utilisation rates, and redundancy.	Week Seven – Eight
Choosing and Refining Modelling Techniques	Build predictive models (Bayesian, Linear Regression, Random Forest, Gradient Boosting, SVR).Refine models by adjusting features and input variables.	Week Nine
Exploration and Evaluation Against State-of-the-Art Methods	Compare the performance of models using Mean Absolute Error (MAE) and interpret results in relation to existing studies on utilisation optimisation.	Week Ten
Analysis of Results	Analyse findings on booking patterns, redundancy, and predictive accuracy. Summarise insights and link back to research objectives. Draft the research paper and adjust based on supervisor feedback.	Week Ten – Twelve
Presentation of Thesis	Finalise dissertation and prepare presentation for submission and defence.	Week Thirteen

Table 9: Research schedule

APPENDICES

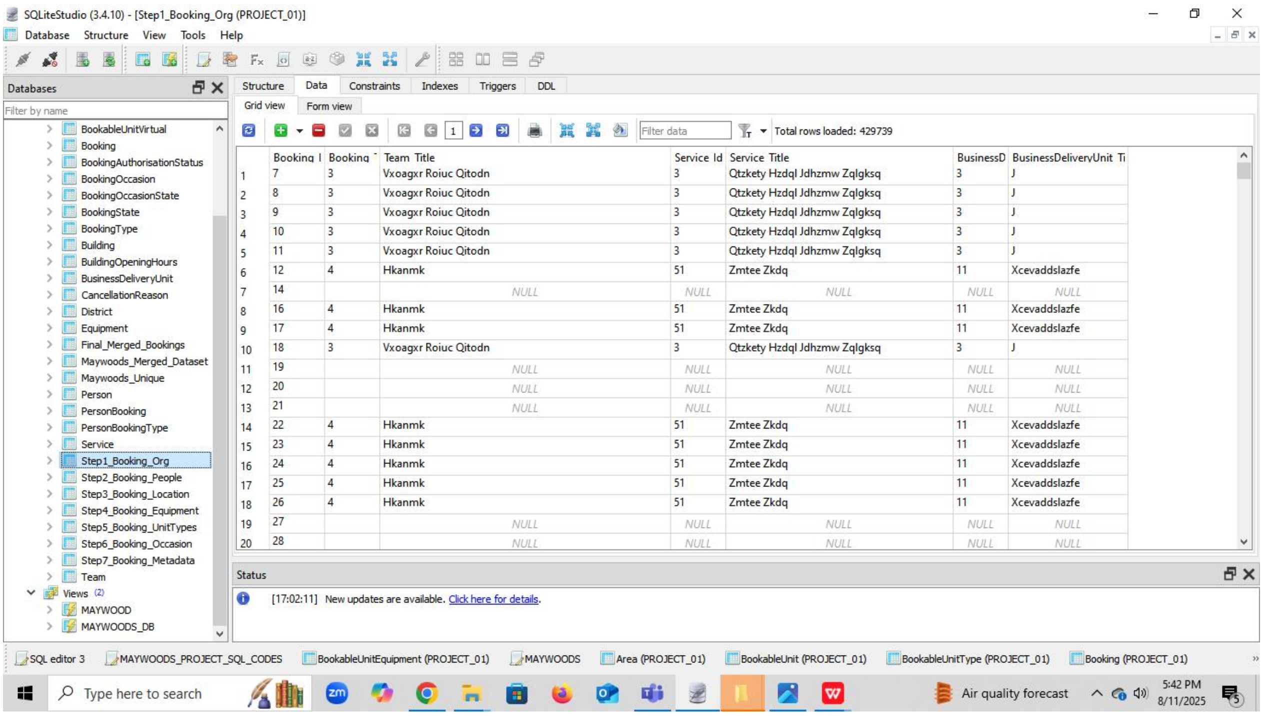
Appendix A: Data Merging in SQL

Step 1



```sql
--STEP 1: Booking + Team + Service + BusinessDeliveryUnit
CREATE TABLE Step1_Booking_Org AS
SELECT
    bk.Booking_Id,
    bk.Booking_TeamId,
    tm.Team_Title,
    srv.Service_Id,
    srv.Service_Title,
    bdu.BusinessDeliveryUnit_Id,
    bdu.BusinessDeliveryUnit_Title
FROM Booking bk
LEFT JOIN Team tm ON bk.Booking_TeamId = tm.Team_Id
LEFT JOIN Service srv ON tm.Team_ServiceId = srv.Service_Id
LEFT JOIN BusinessDeliveryUnit bdu ON srv.Service_BusinessDeliveryUnitId = bdu.BusinessDeliveryUnit_Id;


--STEP 2: PersonBooking + Person + PersonBookingType
CREATE TABLE Step2_Booking_People AS
SELECT
    pb.PersonBooking_Id,
    pb.PersonBooking_BookingId,
    p.Person_Id,
    pbt.PersonBookingType_Id,
    pbt.PersonBookingType_Title
```

Fig. 3.1a(i): SQL merging step 1



Total rows loaded: 429739

	Booking I	Booking ˉ	Team Title	Service Id	Service Title	BusinessD	BusinessDeliveryUnit Ti
1	7	3	Vxoagxr Roiuc Qitodn	3	Qtzkety Hzdql Jdhzmw Zqlgksq	3	J
2	8	3	Vxoagxr Roiuc Qitodn	3	Qtzkety Hzdql Jdhzmw Zqlgksq	3	J
3	9	3	Vxoagxr Roiuc Qitodn	3	Qtzkety Hzdql Jdhzmw Zqlgksq	3	J
4	10	3	Vxoagxr Roiuc Qitodn	3	Qtzkety Hzdql Jdhzmw Zqlgksq	3	J
5	11	3	Vxoagxr Roiuc Qitodn	3	Qtzkety Hzdql Jdhzmw Zqlgksq	3	J
6	12	4	Hkanmk	51	Zmtee Zkdq	11	Xcevaddslazfe
7	14		NULL	NULL	NULL	NULL	NULL
8	16	4	Hkanmk	51	Zmtee Zkdq	11	Xcevaddslazfe
9	17	4	Hkanmk	51	Zmtee Zkdq	11	Xcevaddslazfe
10	18	3	Vxoagxr Roiuc Qitodn	3	Qtzkety Hzdql Jdhzmw Zqlgksq	3	J
11	19		NULL	NULL	NULL	NULL	NULL
12	20		NULL	NULL	NULL	NULL	NULL
13	21		NULL	NULL	NULL	NULL	NULL
14	22	4	Hkanmk	51	Zmtee Zkdq	11	Xcevaddslazfe
15	23	4	Hkanmk	51	Zmtee Zkdq	11	Xcevaddslazfe
16	24	4	Hkanmk	51	Zmtee Zkdq	11	Xcevaddslazfe
17	25	4	Hkanmk	51	Zmtee Zkdq	11	Xcevaddslazfe
18	26	4	Hkanmk	51	Zmtee Zkdq	11	Xcevaddslazfe
19	27		NULL	NULL	NULL	NULL	NULL
20	28		NULL	NULL	NULL	NULL	NULL

Fig. 3.1a(ii): SQL merging step 1

Step 2

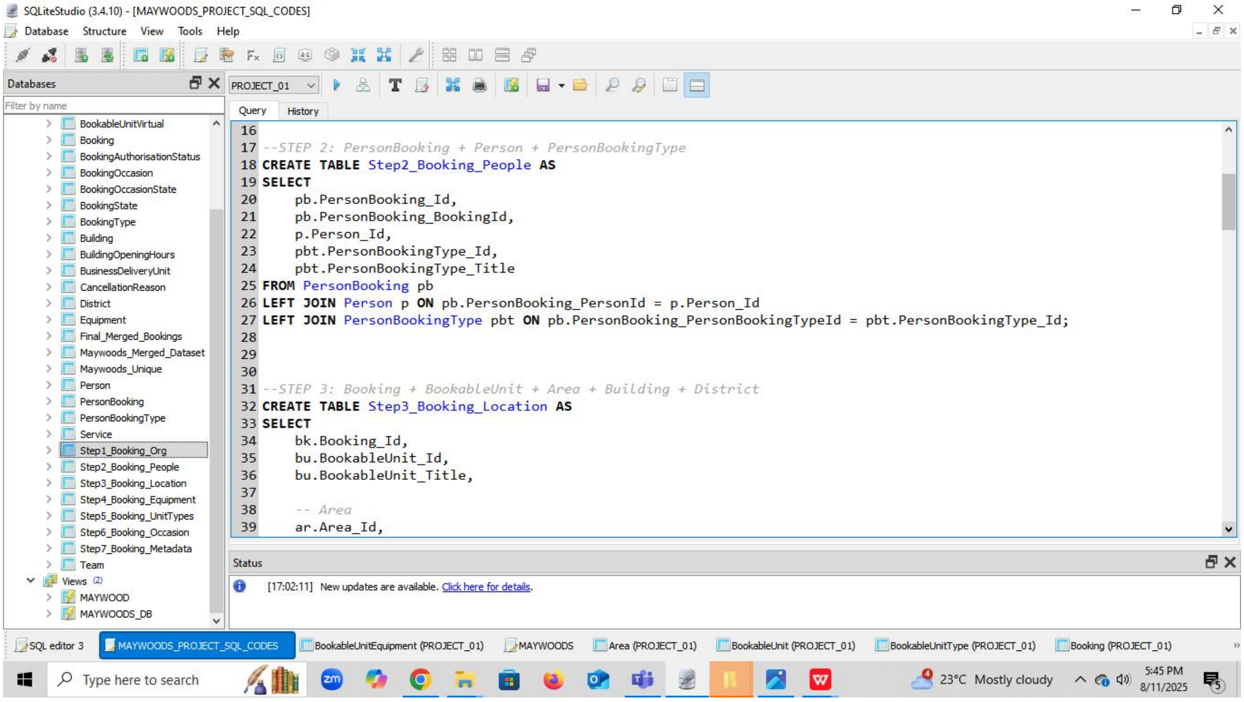


Fig. 3.1b(i): SQL merging step 2

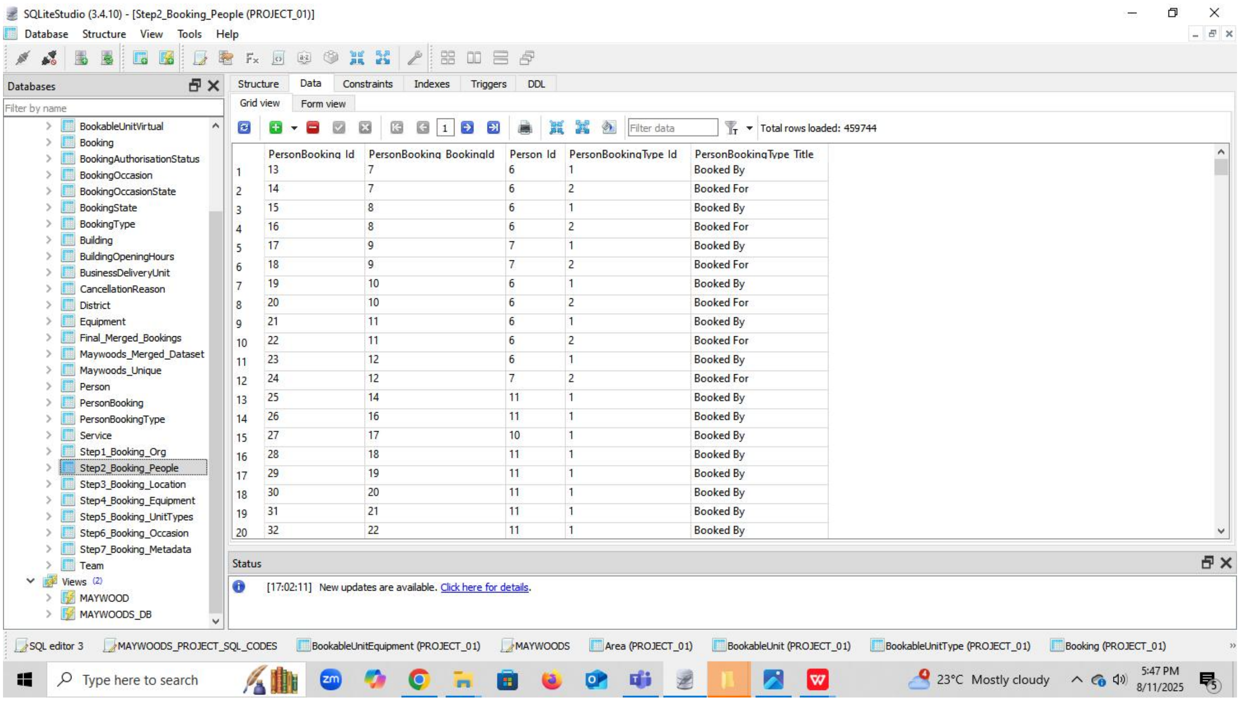


Fig. 3.1b(ii): SQL merging step 2

Step 3

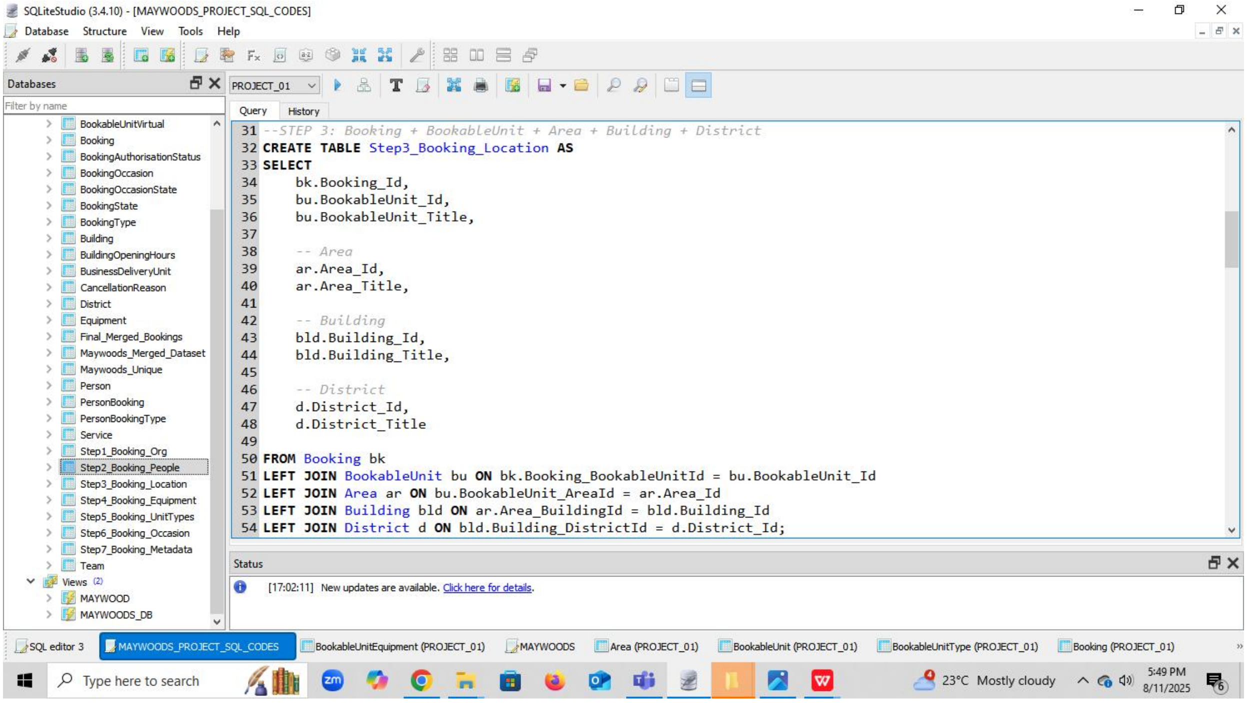


Fig. 3.1c(i): SQL merging step 3

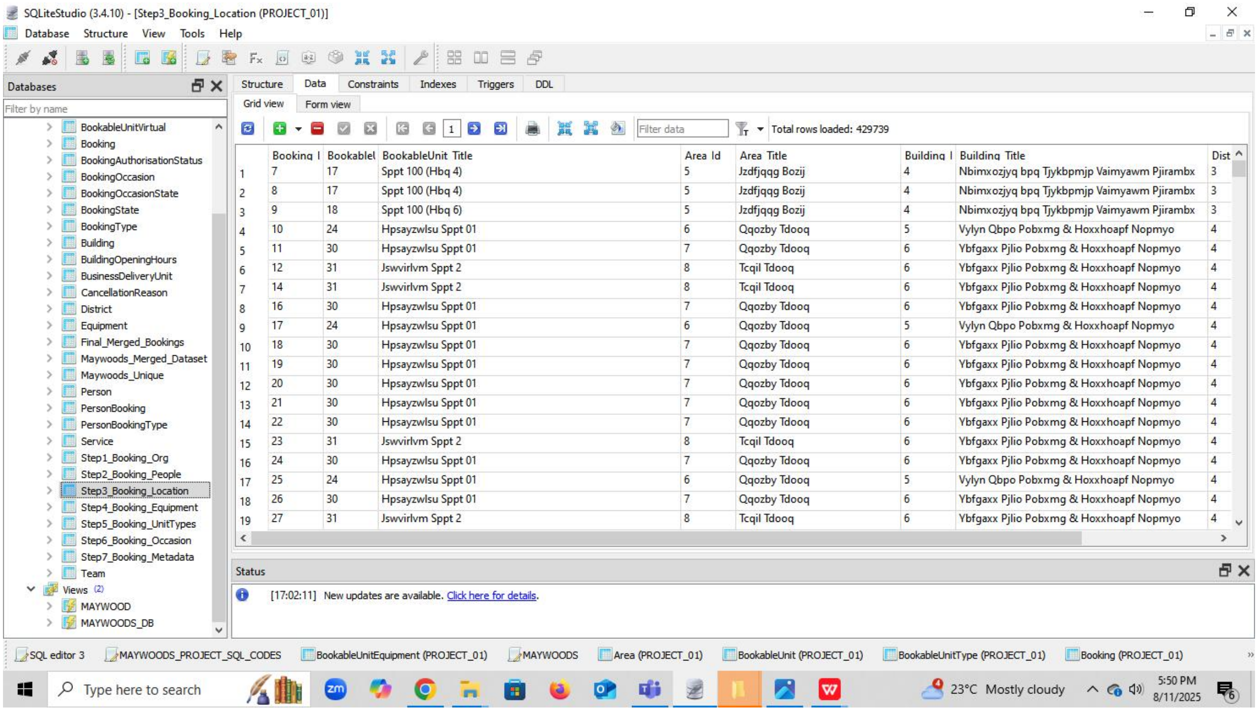


Fig. 3.1c(ii): SQL merging step 3

Step 4

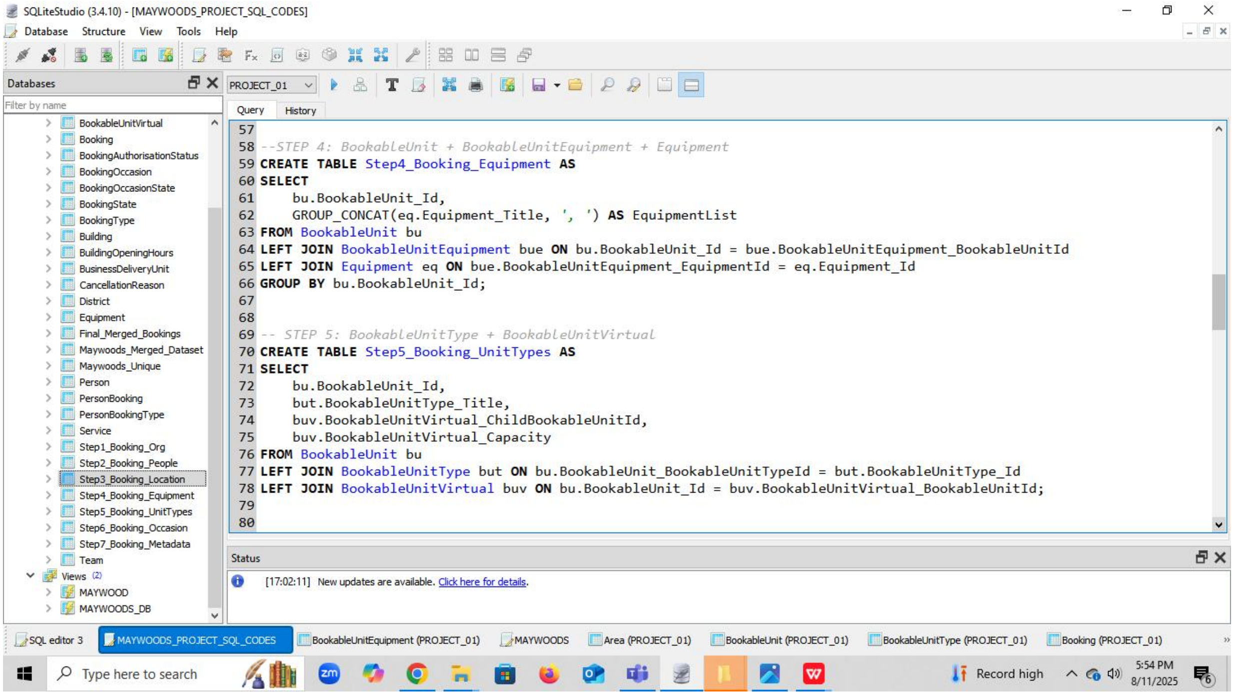


Fig. 3.1d(i): SQL merging step 4

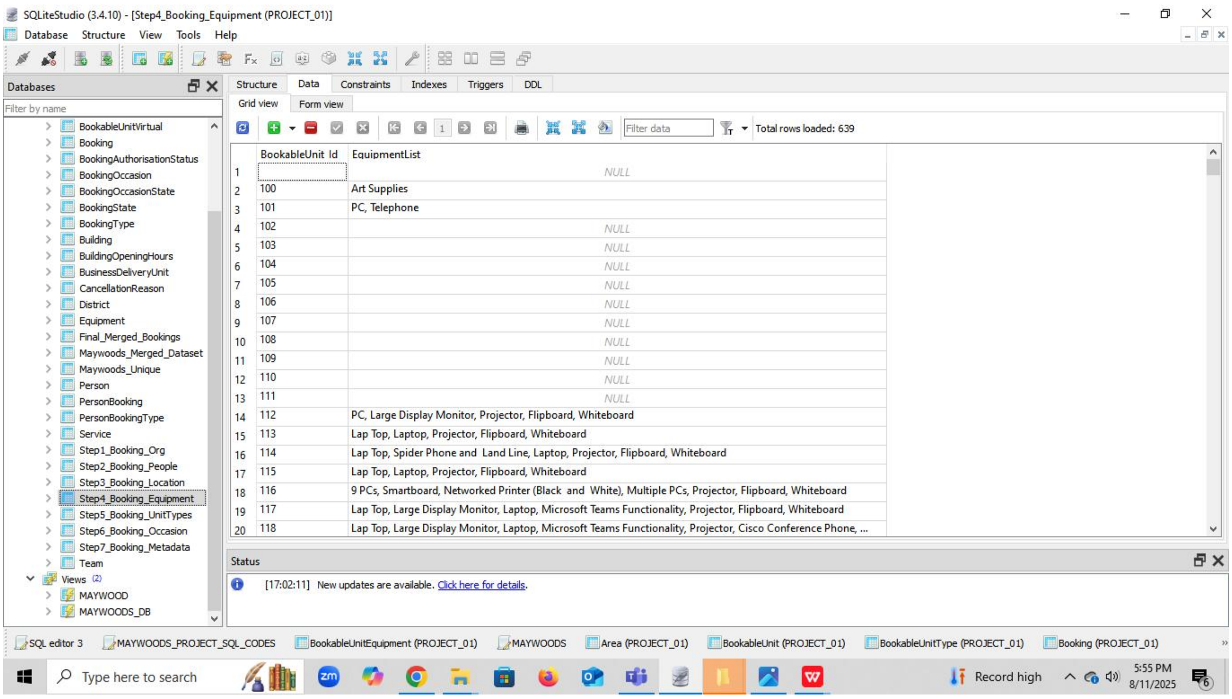


Fig. 3.1d(ii): SQL merging step 4

Step 5

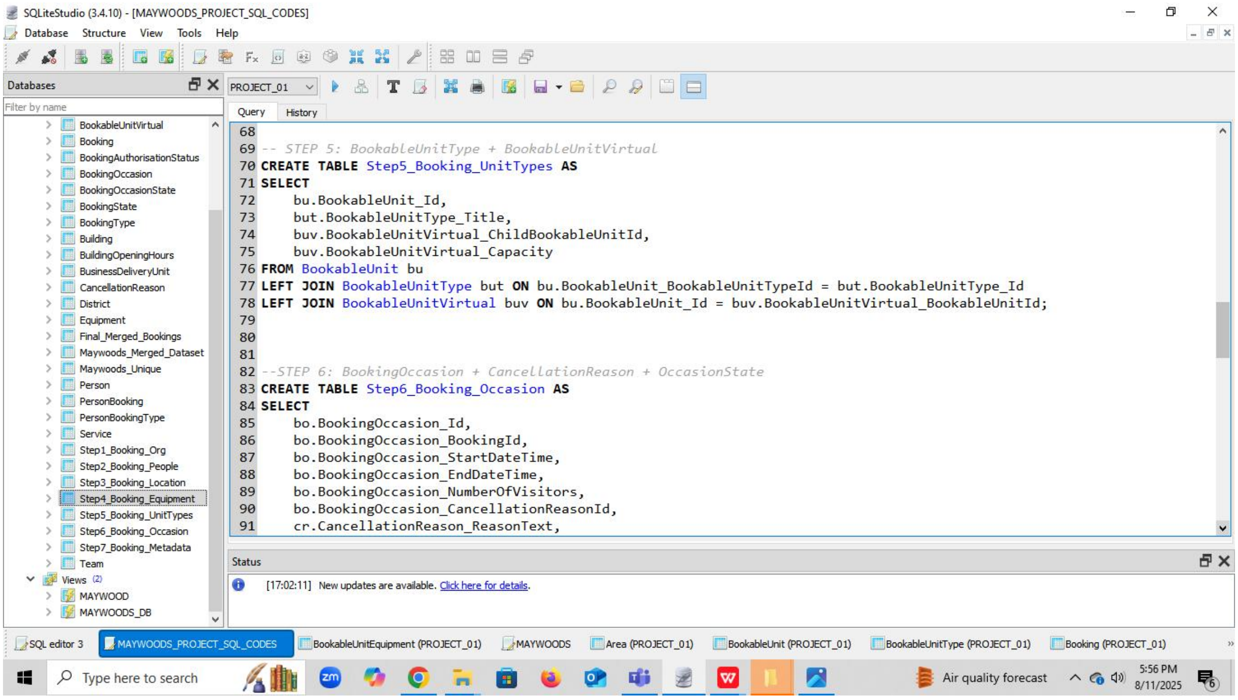


Fig. 3.1e(i): SQL merging step 5

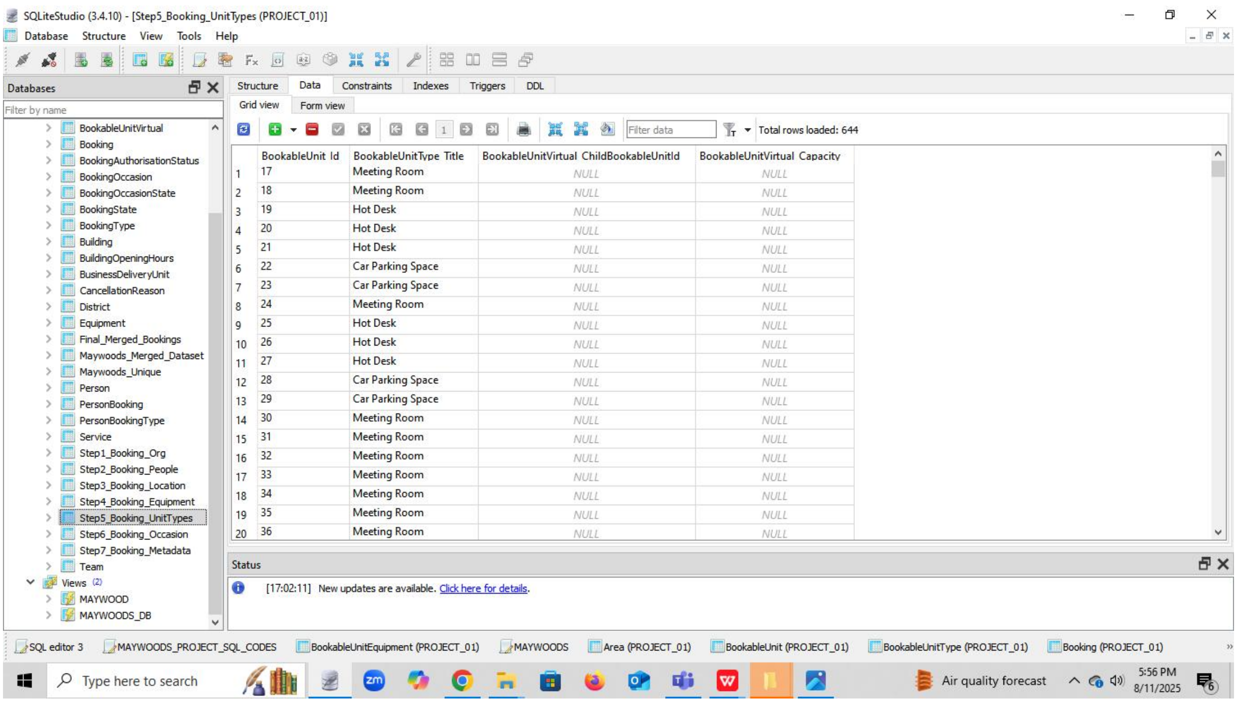


Fig. 3.1e(ii): SQL merging step 5

Step 6

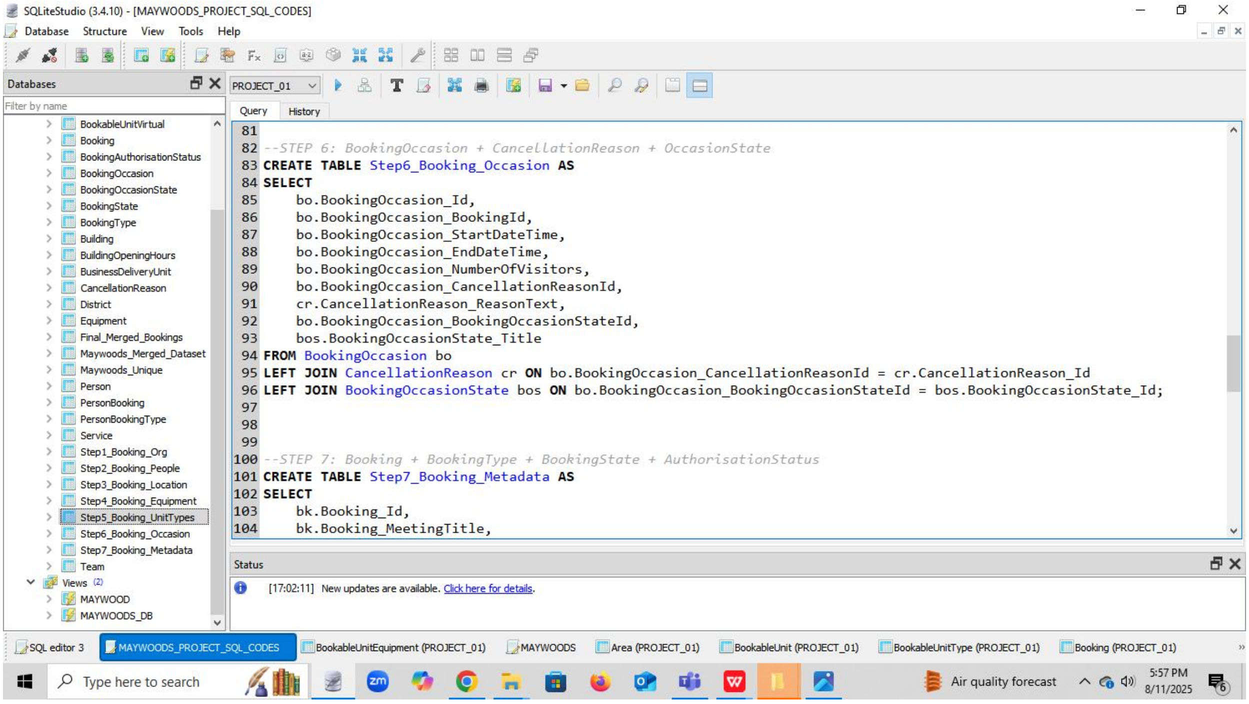


Fig. 3.1f(i): SQL merging step 6

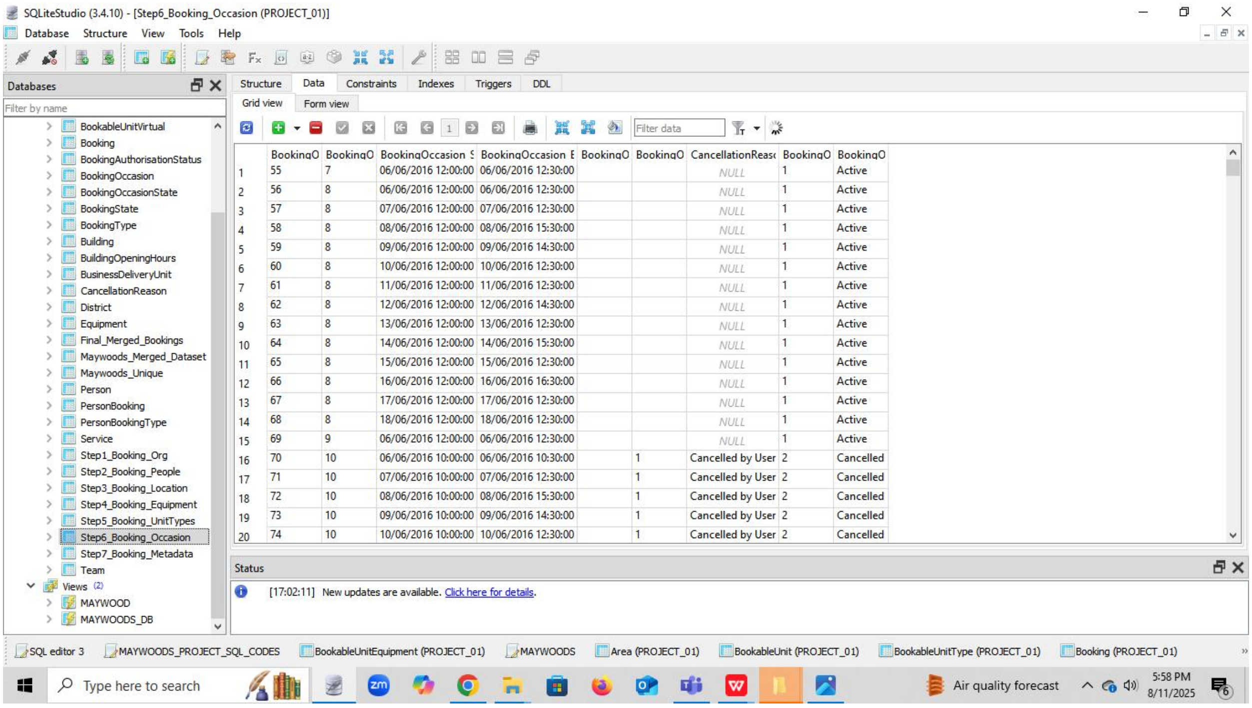


Fig. 3.1f(ii): SQL merging step 6

Step 7

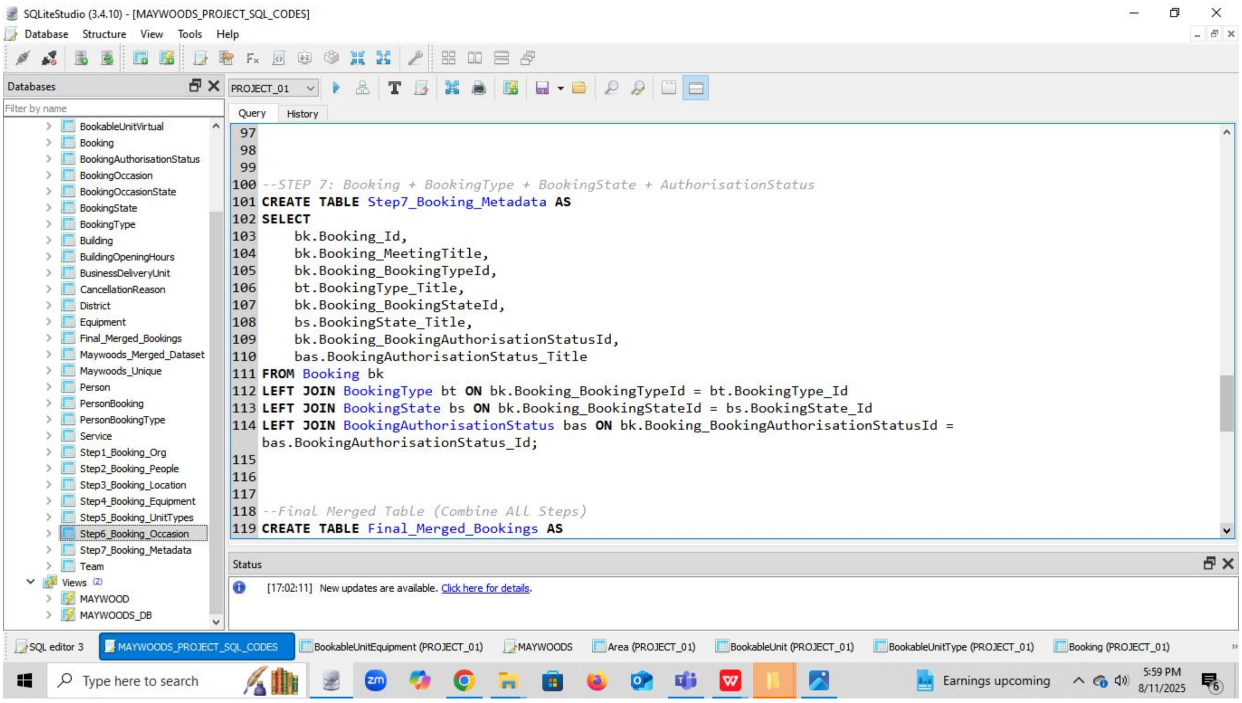


Fig. 3.1g(i): SQL merging step 7

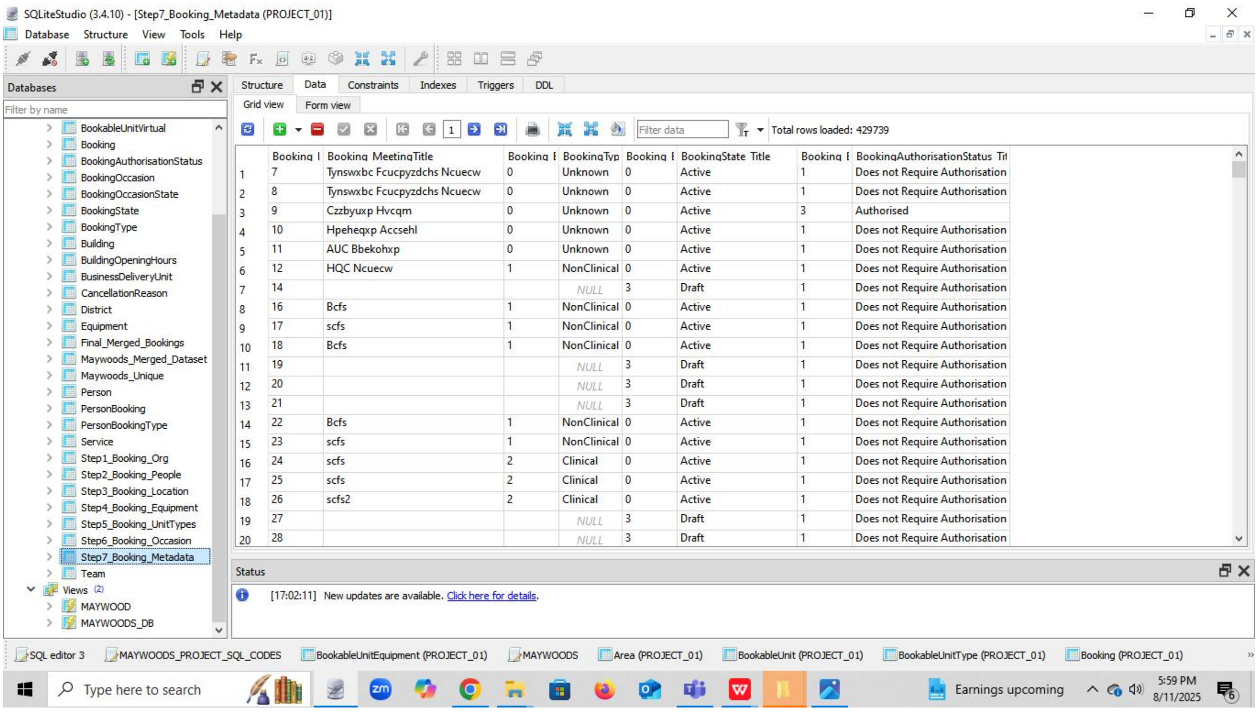


Fig. 3.1g(ii): SQL merging step 7

Merging of all steps

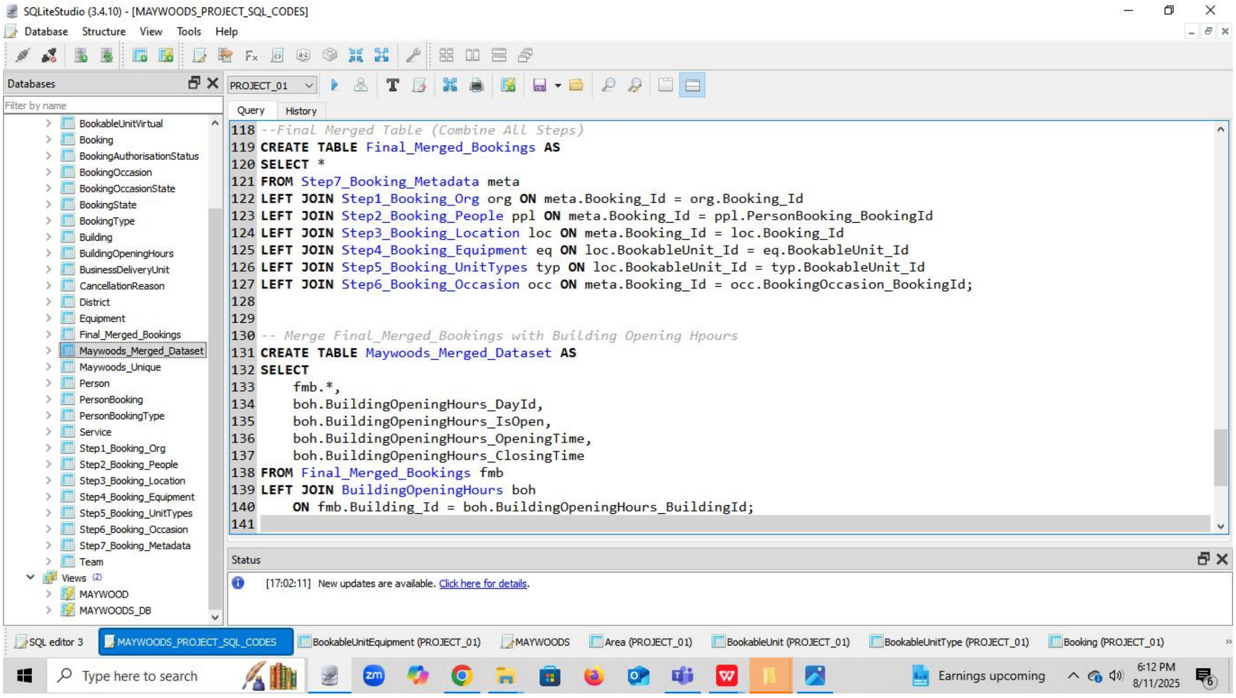


Fig. 3.1h(i): SQL merging of all steps

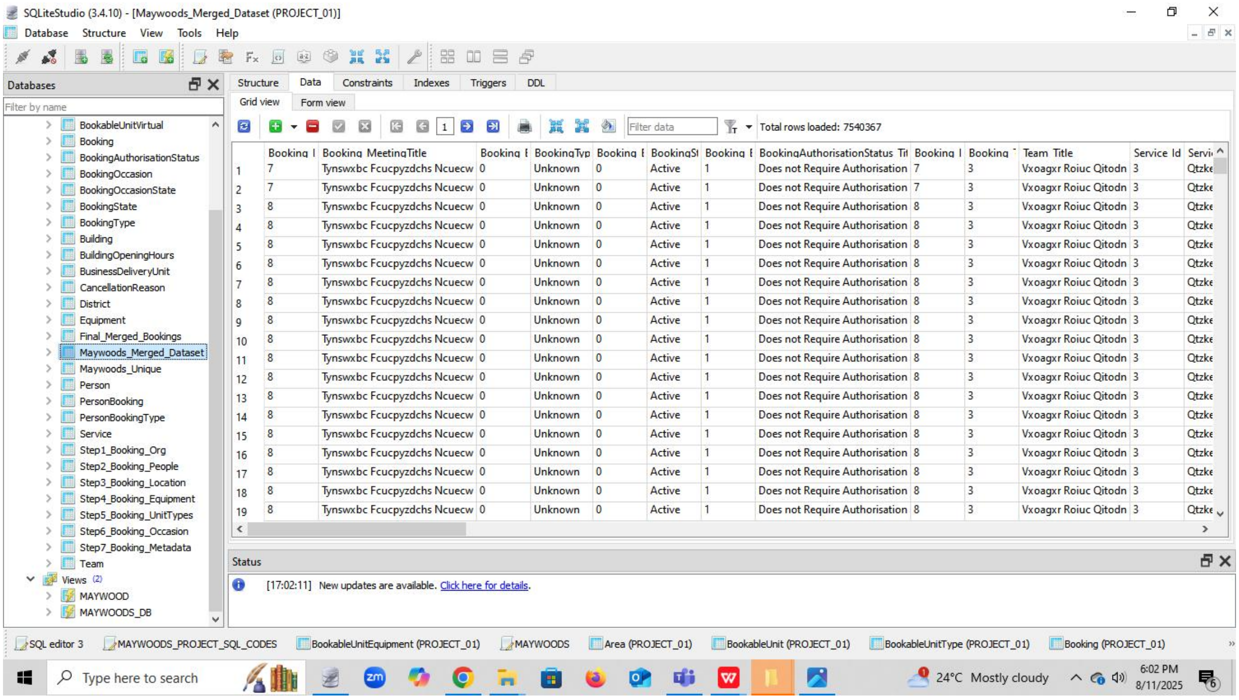


Fig. 3.1h(ii): SQL merging of all steps

Drop duplicated rows

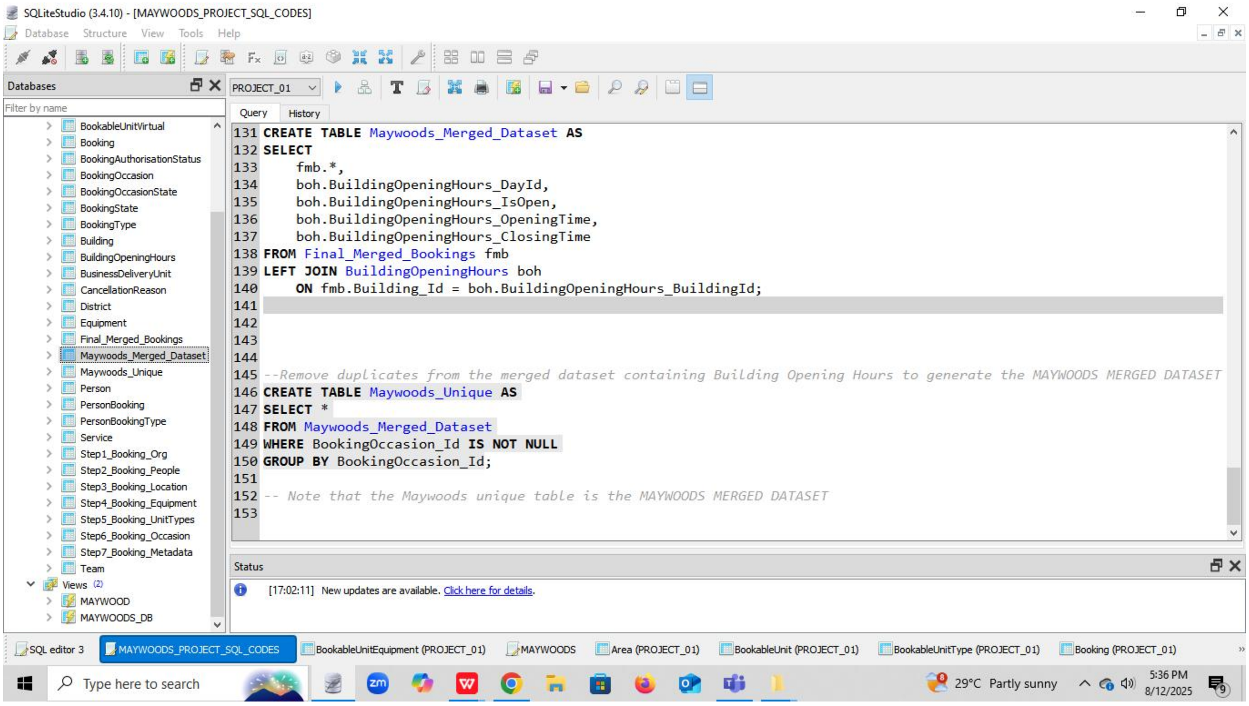


Fig. 3.1h(iii): Dropping of duplicated rows

Appendix B: Data Preprocessing

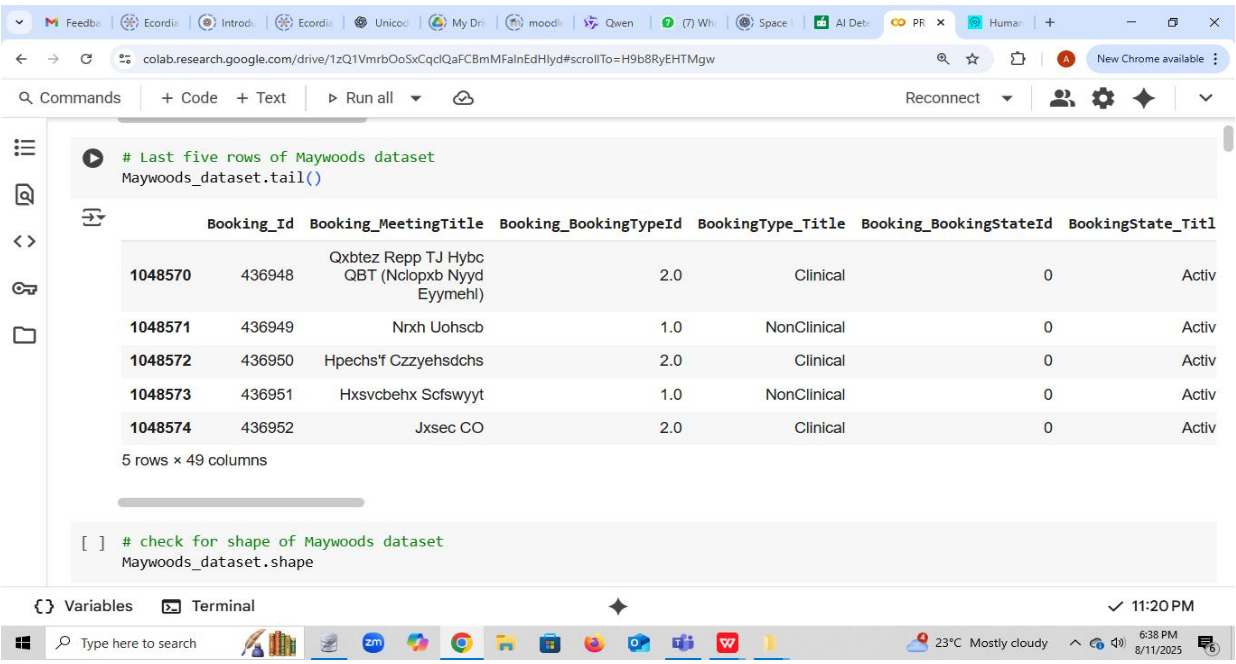


Fig. 5.3: Last 5 rows of Maywoods dataset

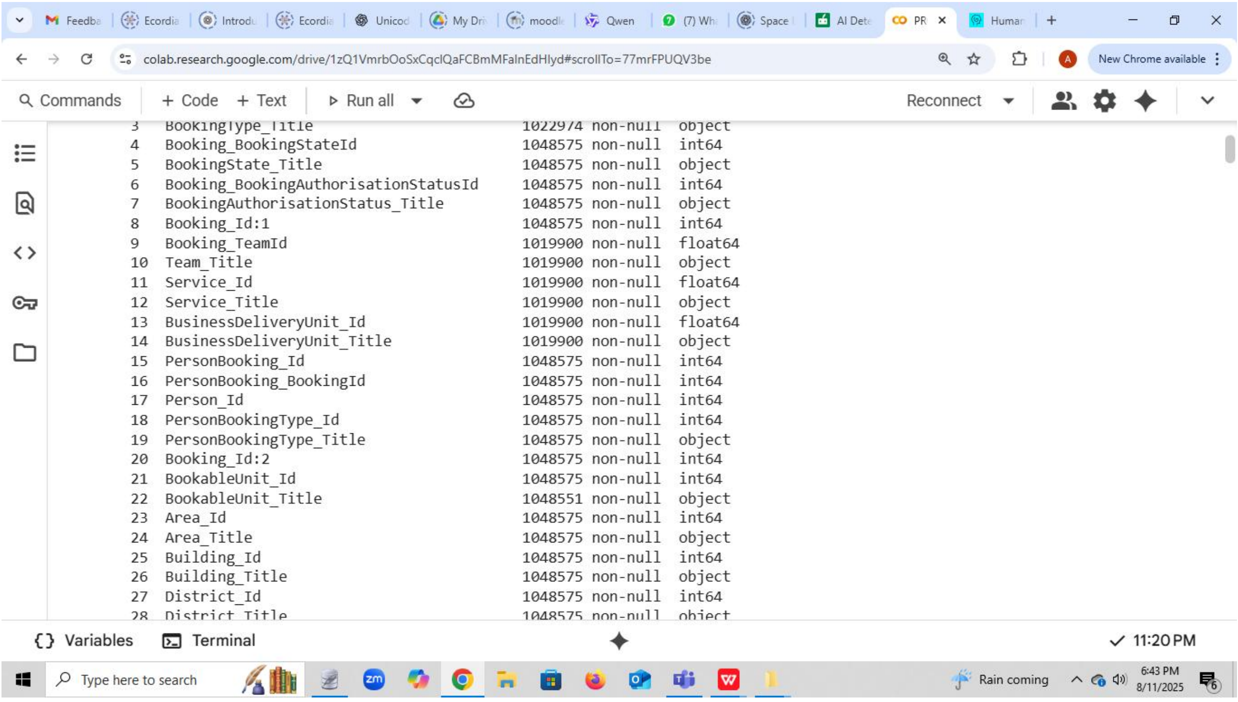


Fig. 6.2: Maywoods dataset info 2

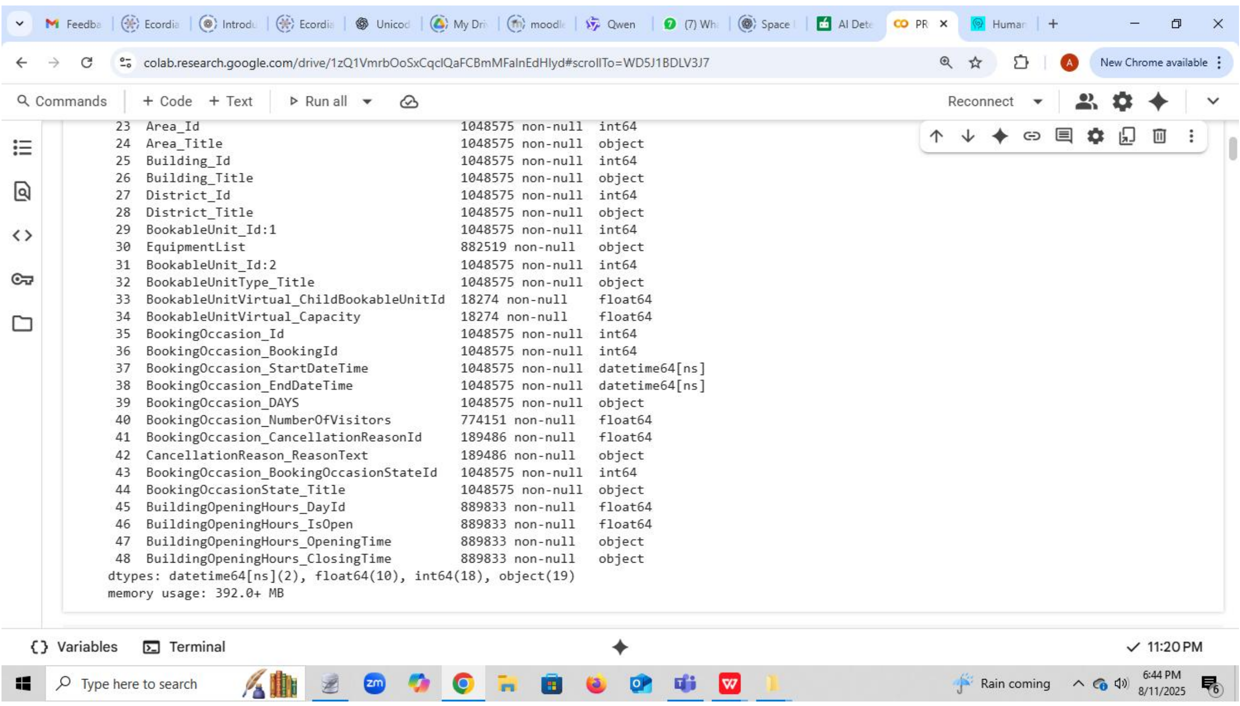
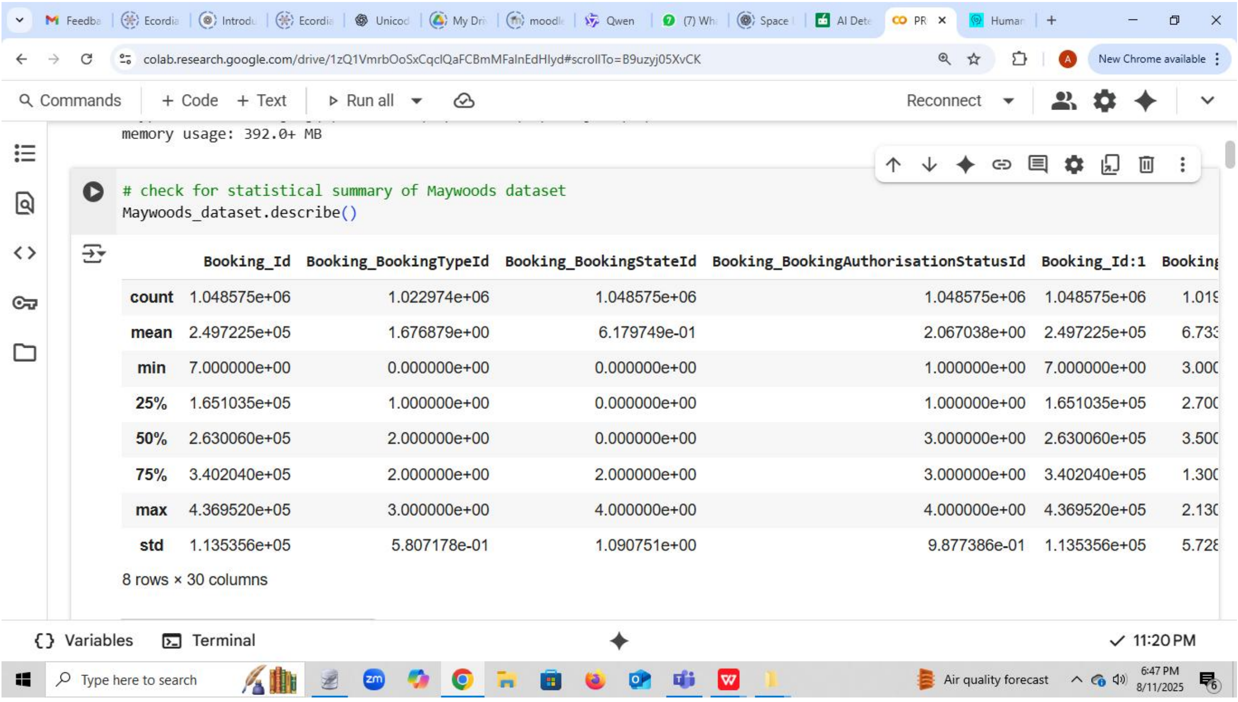


Fig. 6.3: Maywoods dataset info 3

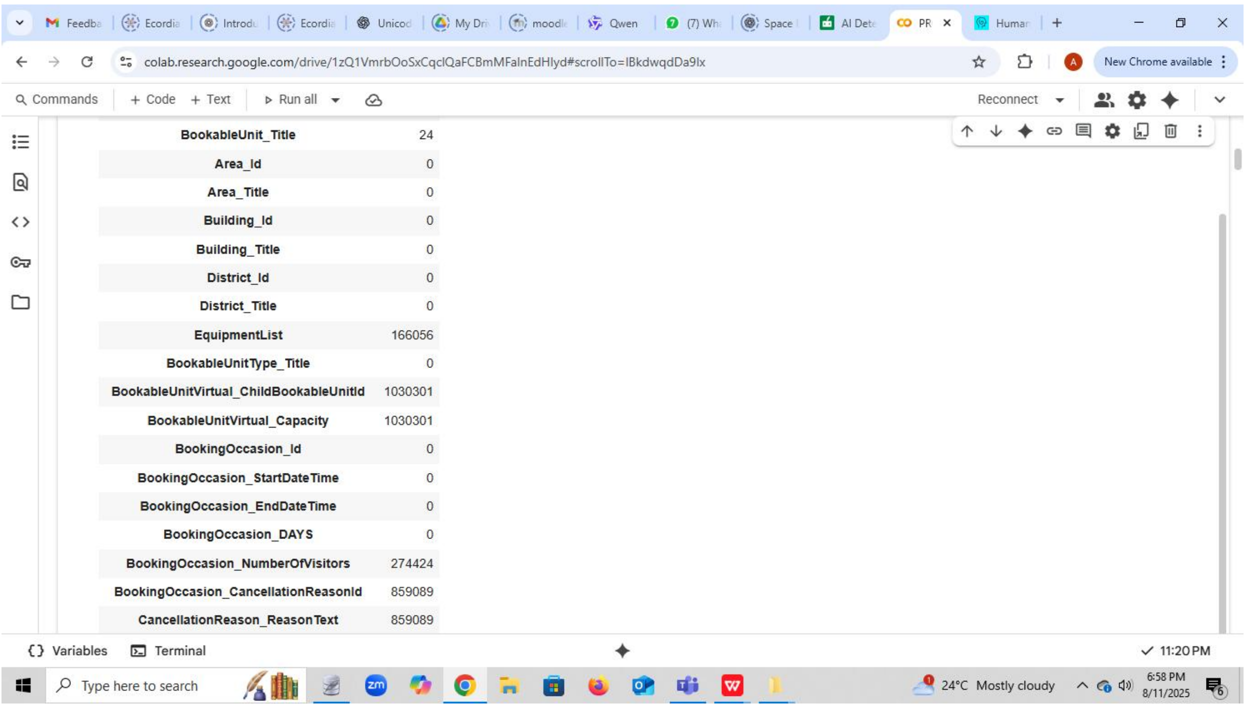


```
# check for statistical summary of Maywoods dataset
Maywoods_dataset.describe()
```

	Booking_Id	Booking_BookingTypeId	Booking_BookingStateId	Booking_BookingAuthorisationStatusId	Booking_Id:1	Booking
count	1.048575e+06	1.022974e+06	1.048575e+06	1.048575e+06	1.048575e+06	1.019
mean	2.497225e+05	1.676879e+00	6.179749e-01	2.067038e+00	2.497225e+05	6.733
min	7.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00	7.000000e+00	3.000
25%	1.651035e+05	1.000000e+00	0.000000e+00	1.000000e+00	1.651035e+05	2.700
50%	2.630060e+05	2.000000e+00	0.000000e+00	3.000000e+00	2.630060e+05	3.500
75%	3.402040e+05	2.000000e+00	2.000000e+00	3.000000e+00	3.402040e+05	1.300
max	4.369520e+05	3.000000e+00	4.000000e+00	4.000000e+00	4.369520e+05	2.130
std	1.135356e+05	5.807178e-01	1.090751e+00	9.877386e-01	1.135356e+05	5.728

8 rows × 30 columns

Fig. 6.4: Statistical summary of Maywoods dataset



BookableUnit_Title	24
Area_Id	0
Area_Title	0
Building_Id	0
Building_Title	0
District_Id	0
District_Title	0
EquipmentList	166056
BookableUnitType_Title	0
BookableUnitVirtual_ChildBookableUnitId	1030301
BookableUnitVirtual_Capacity	1030301
BookingOccasion_Id	0
BookingOccasion_StartDateTime	0
BookingOccasion_EndDateTime	0
BookingOccasion_DAYS	0
BookingOccasion_NumberOfVisitors	274424
BookingOccasion_CancellationReasonId	859089
CancellationReason_ReasonText	859089

Fig. 8.2: Check for null values_2

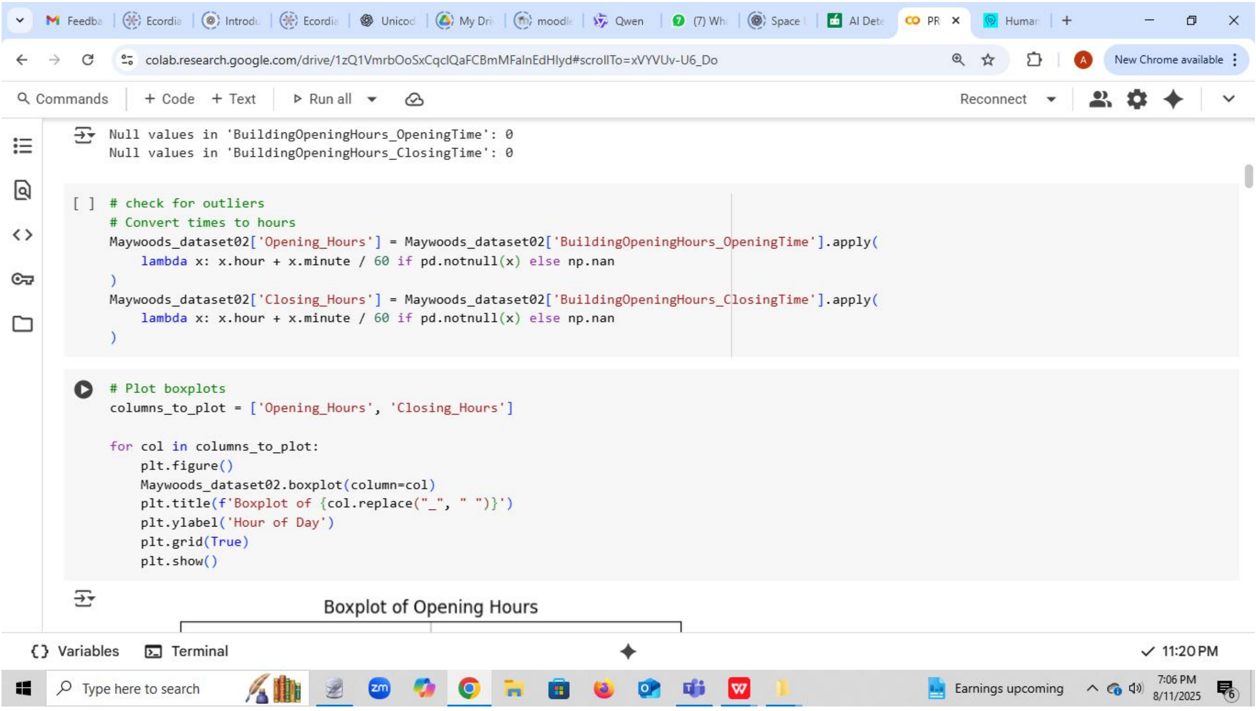


Fig. 11a: Check for outliers

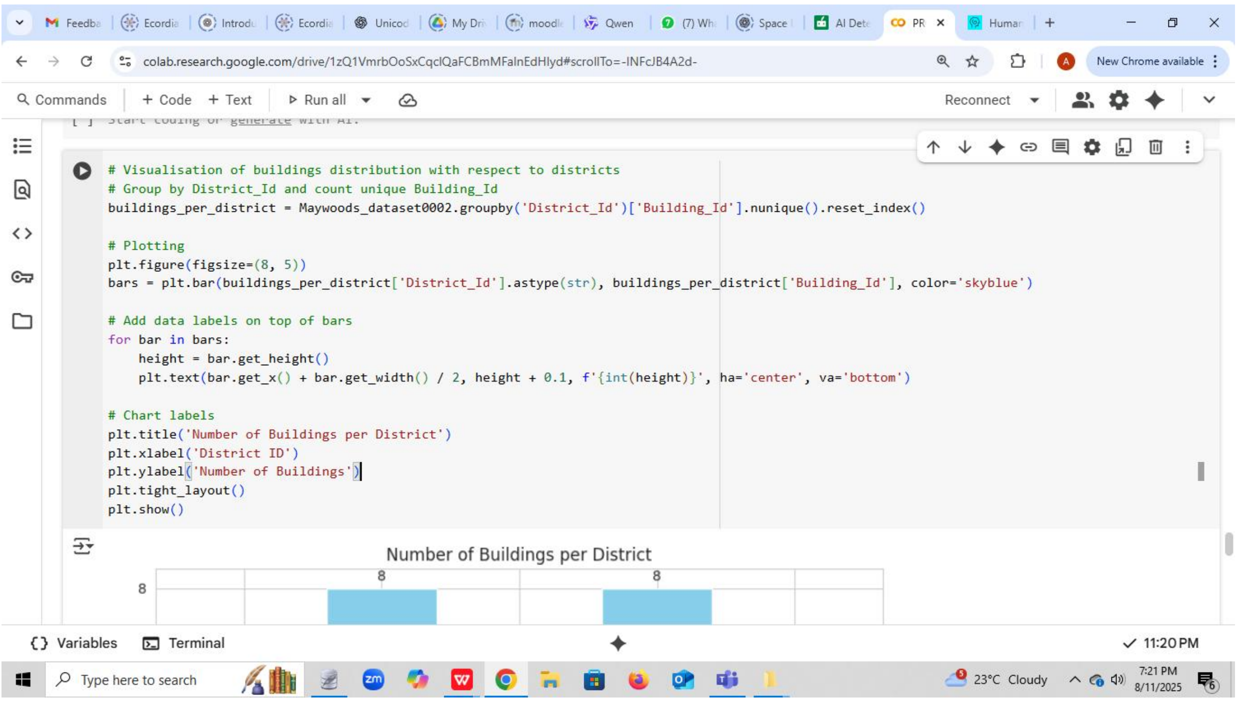


Fig. 12.2(a): Visualisation of buildings distribution with respect to districts 1

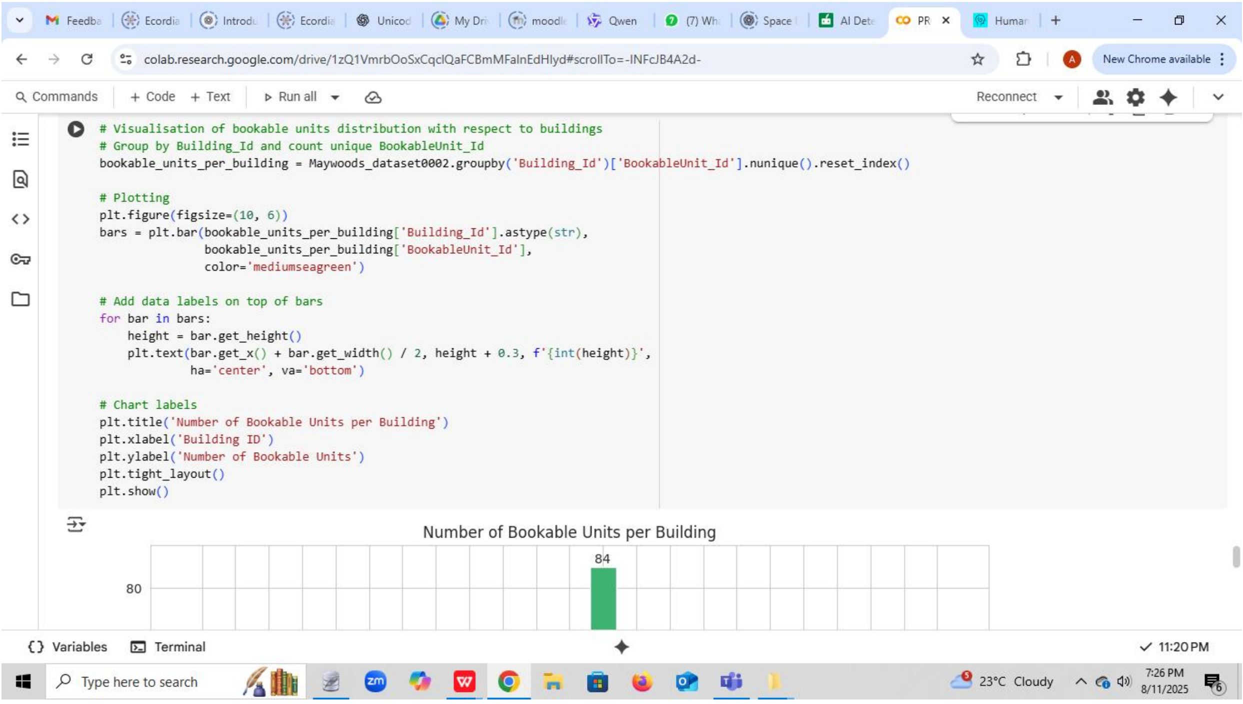


Fig. 12.3(a): Visualisation of bookable units distribution with respect to buildings 1

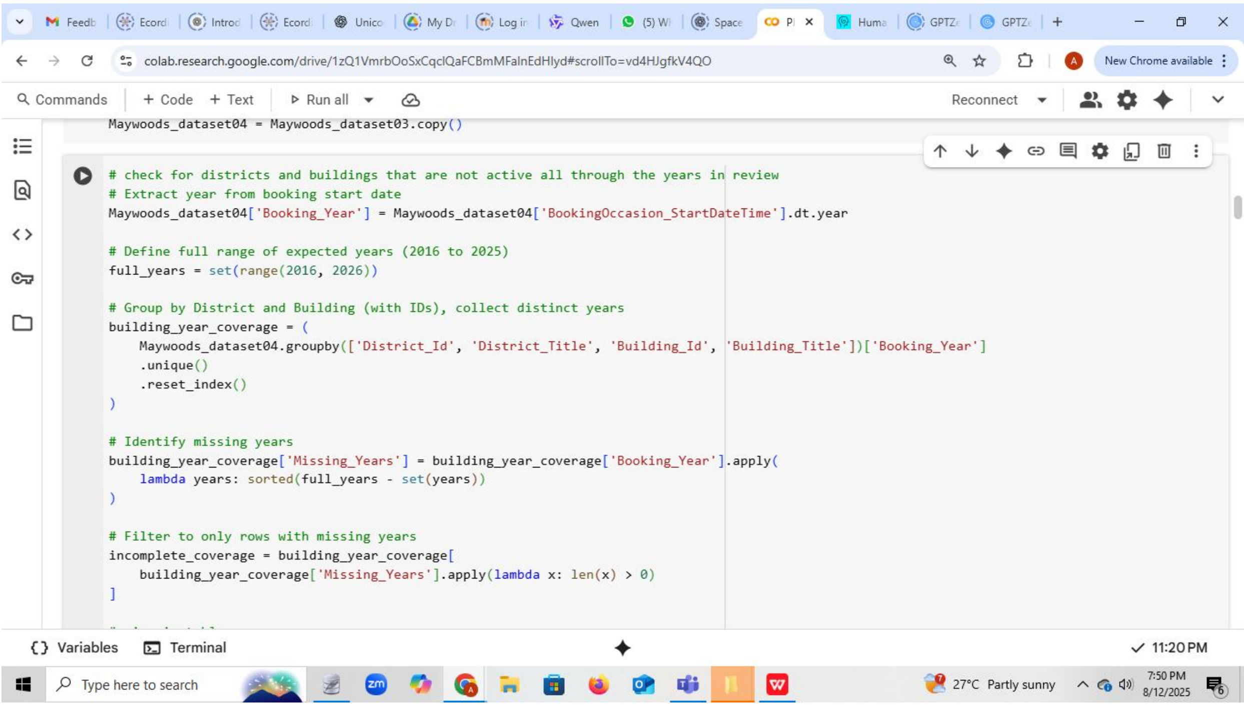


Fig. 14.1: Check for inconsistent districts 1

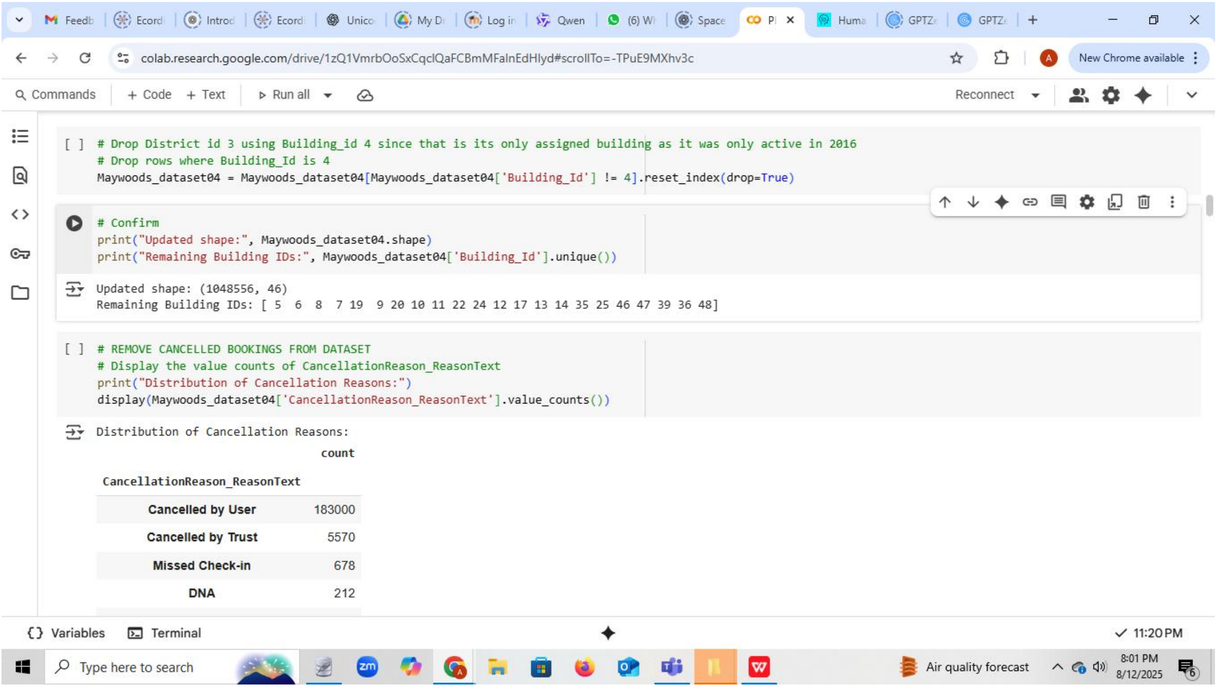


Fig. 14.3: Drop inactive district 3

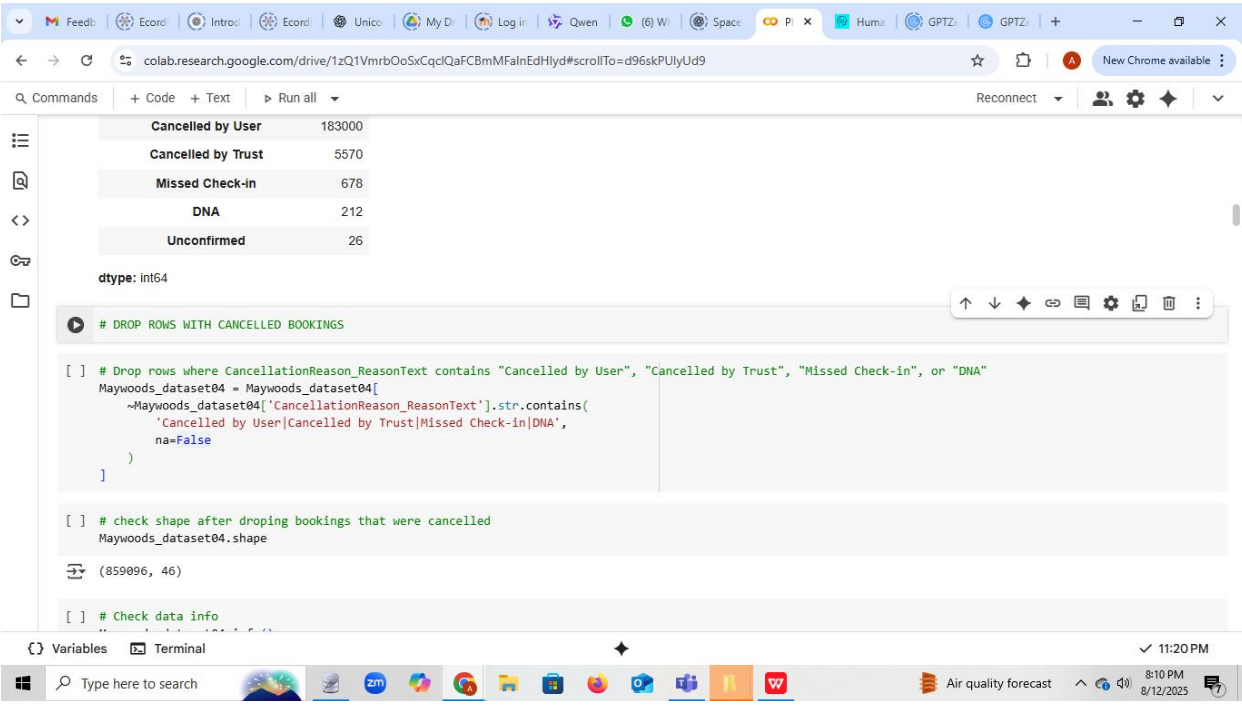


Fig. 14.4: Drop canceled bookings and check dataset shape

Appendix C: Exploratory Data Analysis

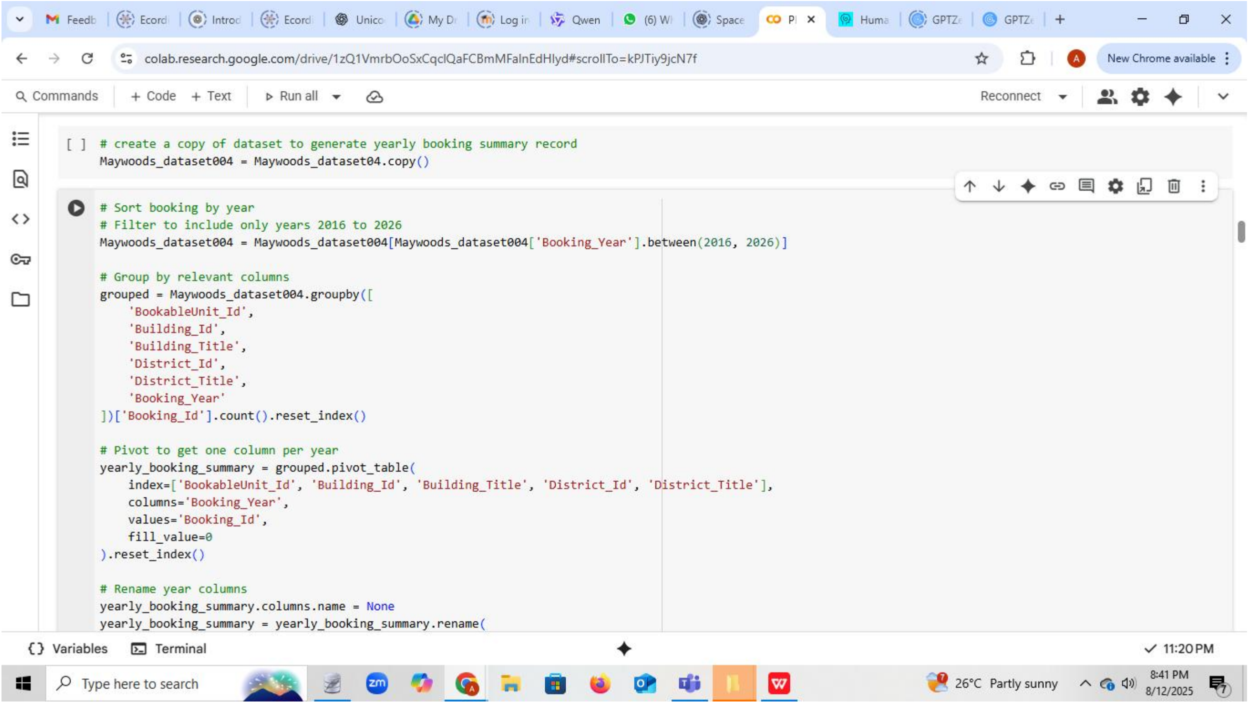


Fig. 16a: Yearly booking summary_1

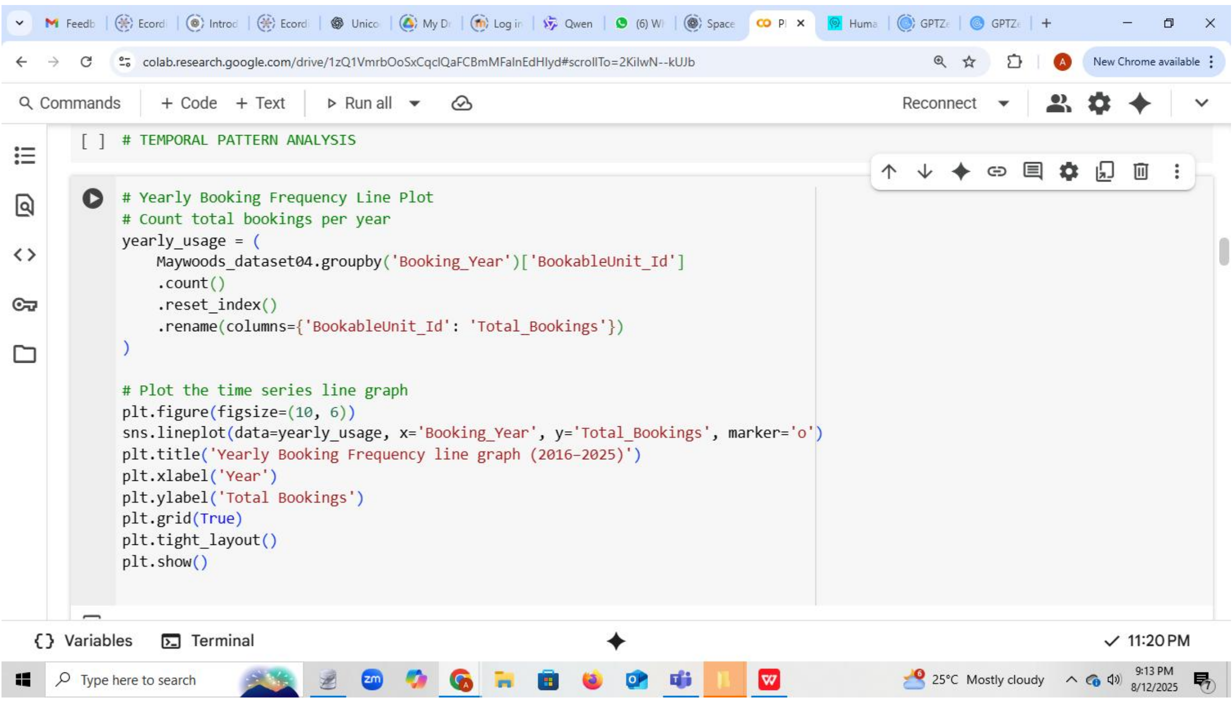


Fig. 17.1: Yearly booking line plot_1

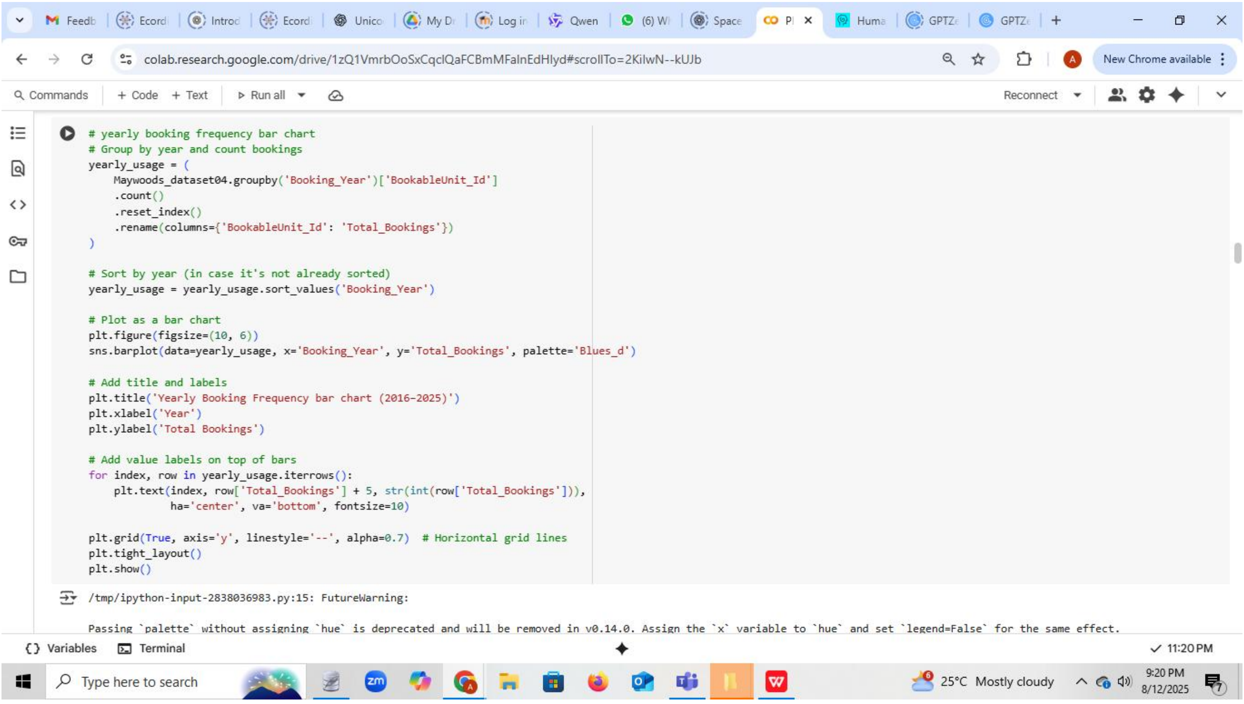


Fig. 17.3: Yearly booking bar chart_1

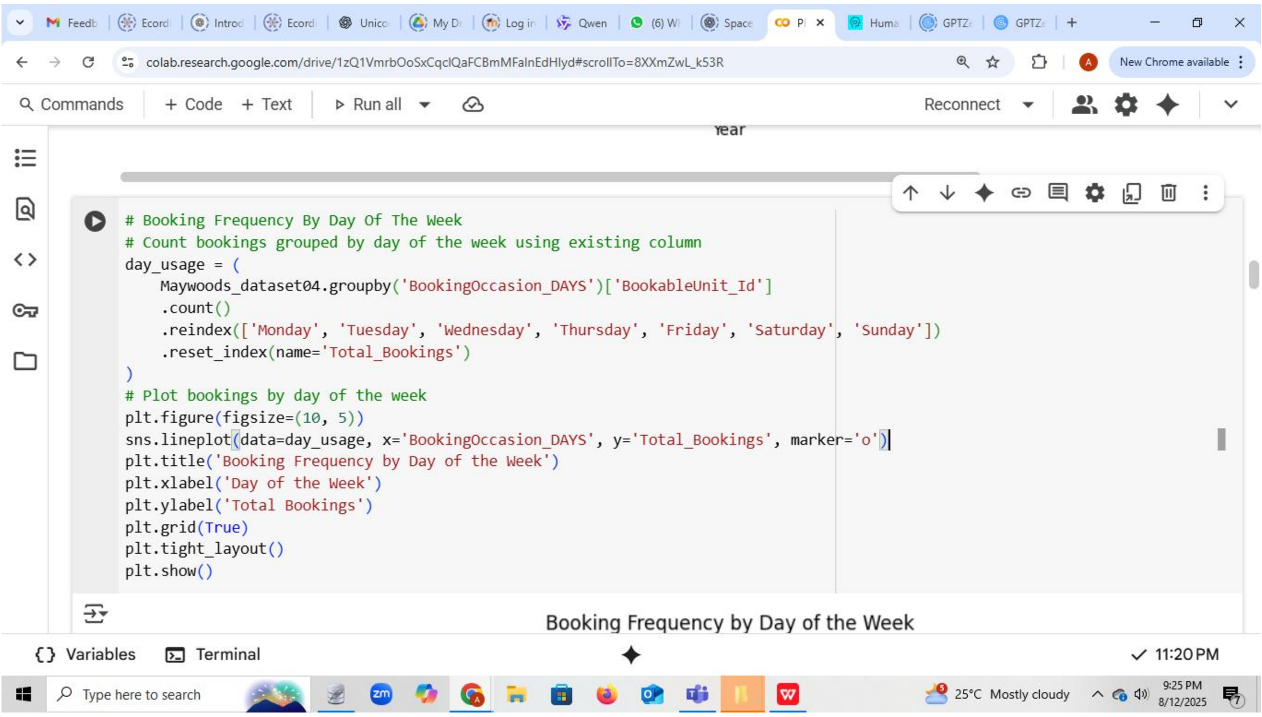


Fig. 18.1: Booking Frequency By Day Of The Week_1

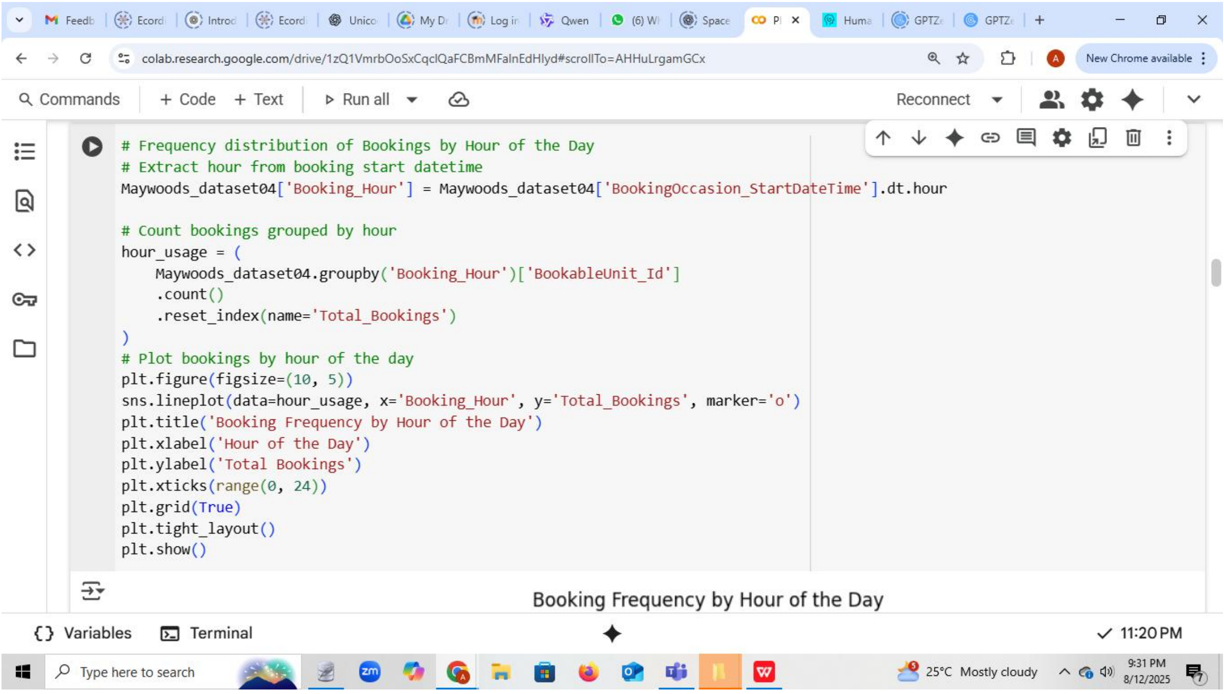


Fig. 19.1: Frequency distribution of Bookings by Hour of the Day_1

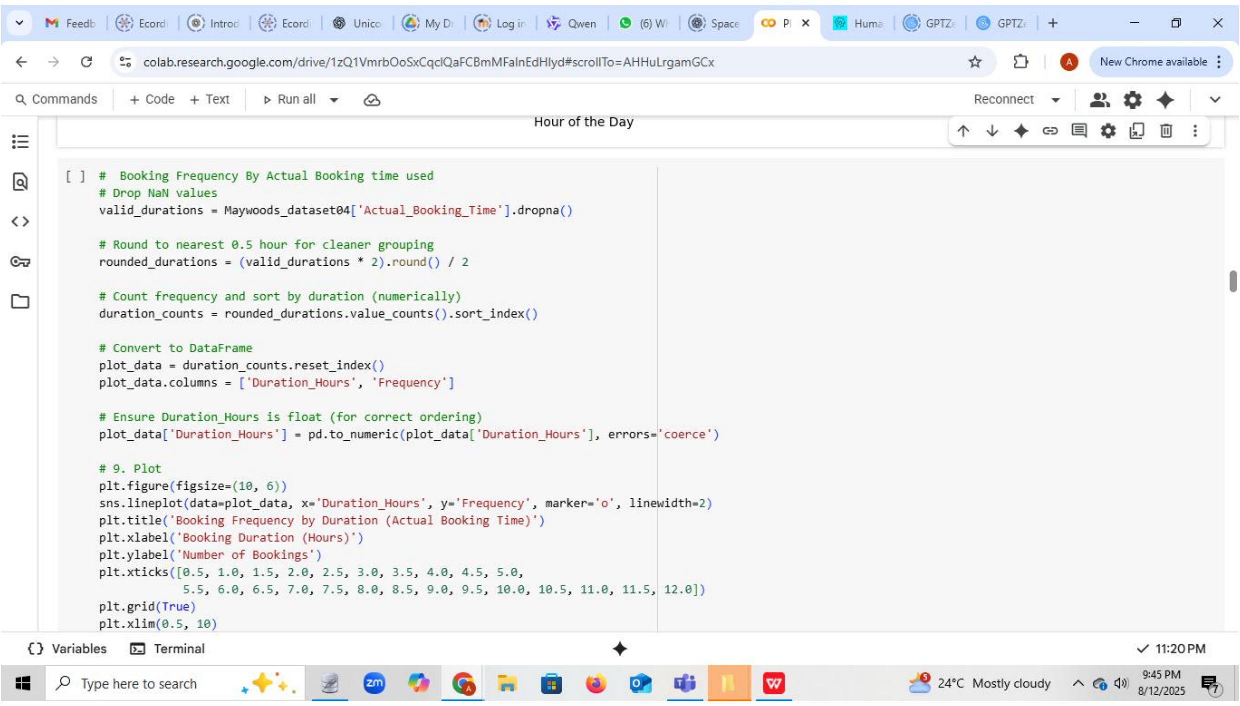


Fig. 20.1: Booking Frequency By Actual Booking time used_1

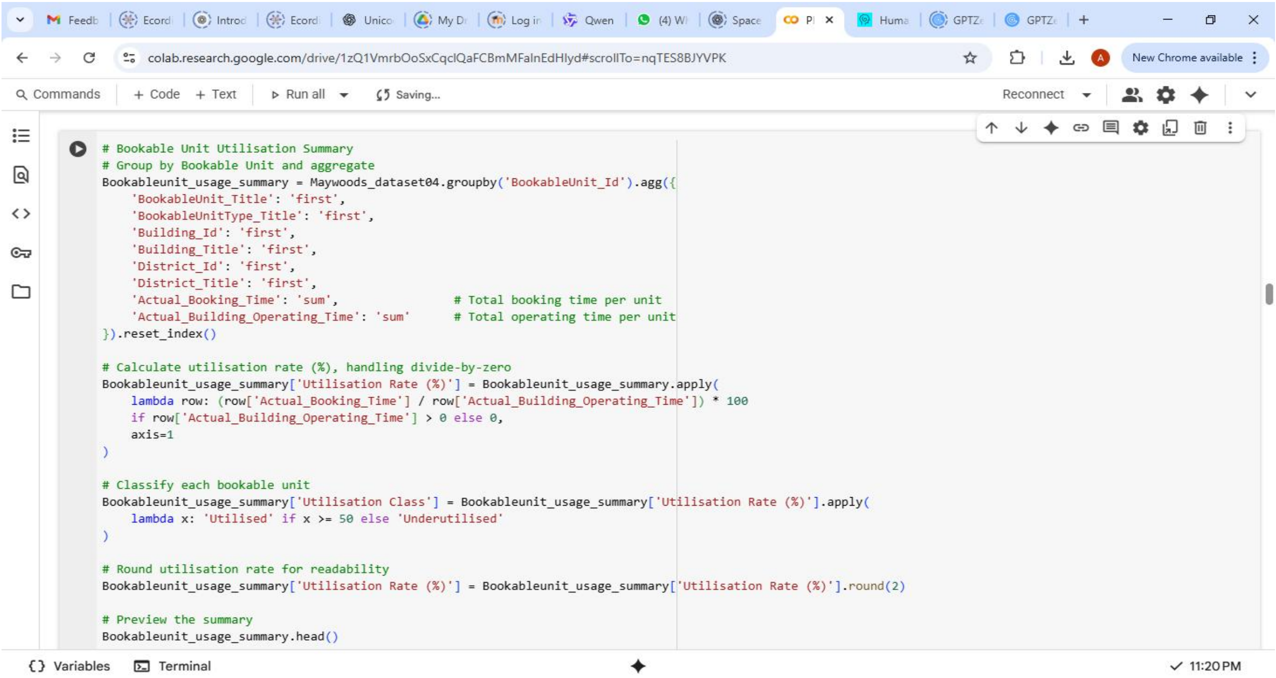


Fig. 21.1: Bookable unit utilisation summary_1

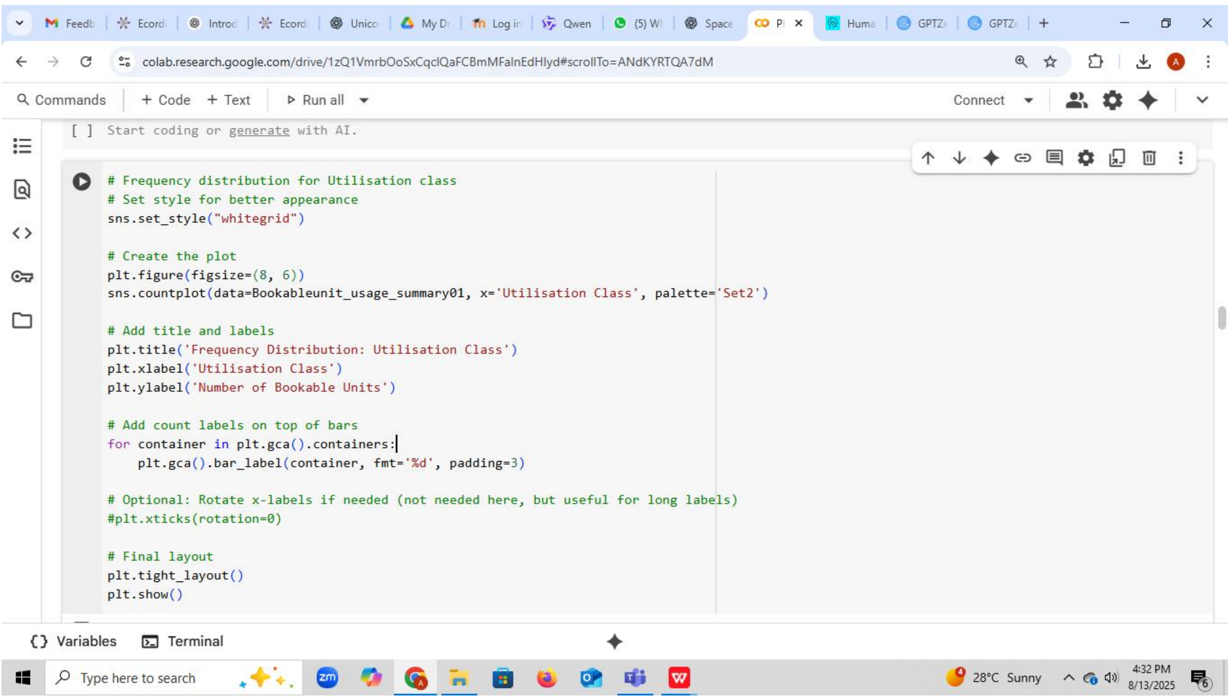


Fig. 22.1: Distribution for utilisation class_1

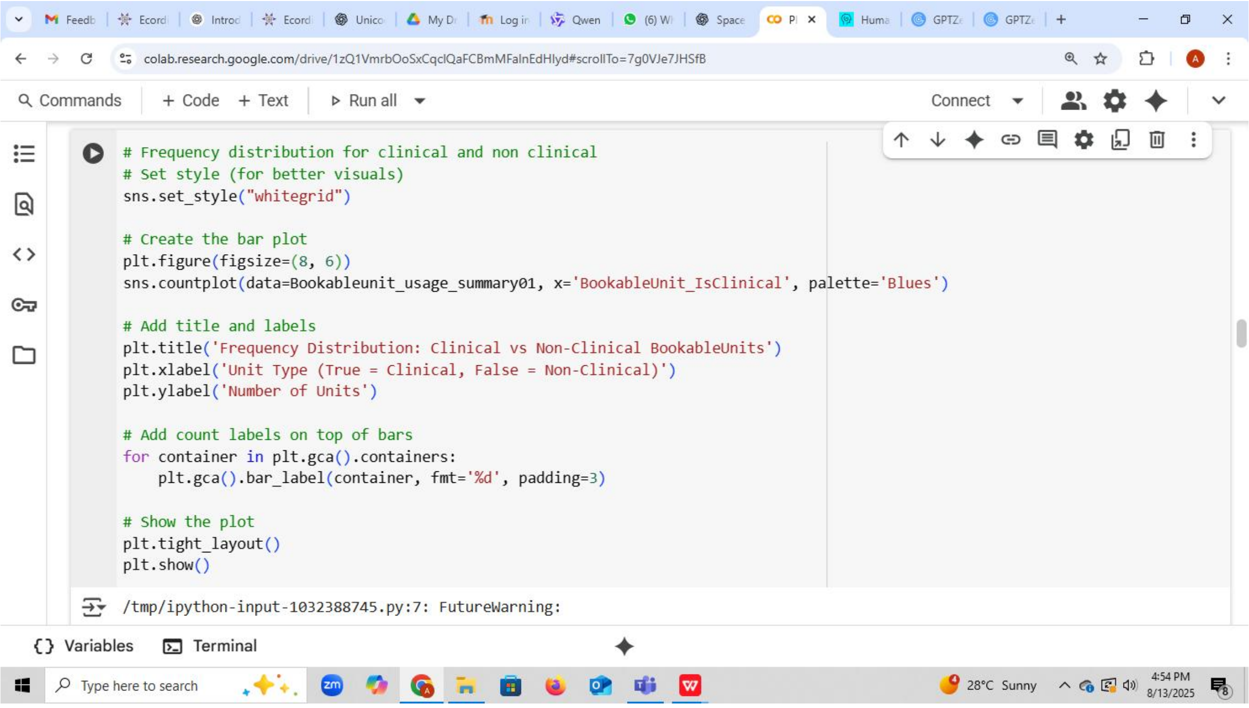


Fig. 23.1: Frequency distribution for clinical and non clinical_1

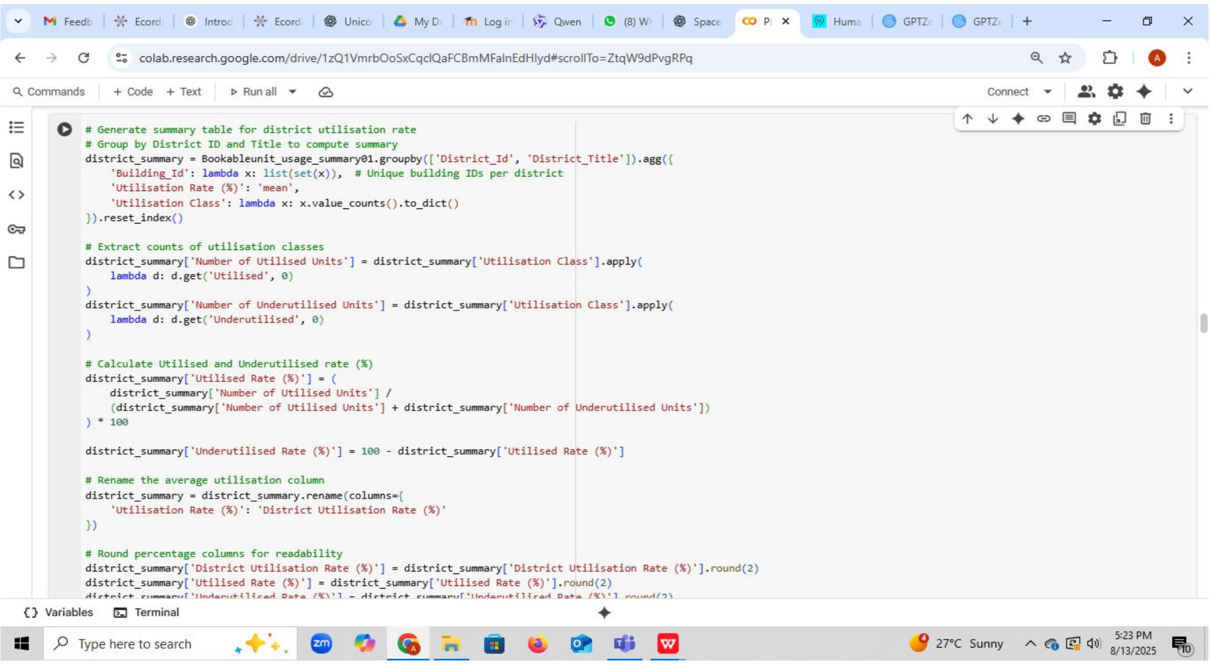


Fig. 24.1: District utilisation rate table_1

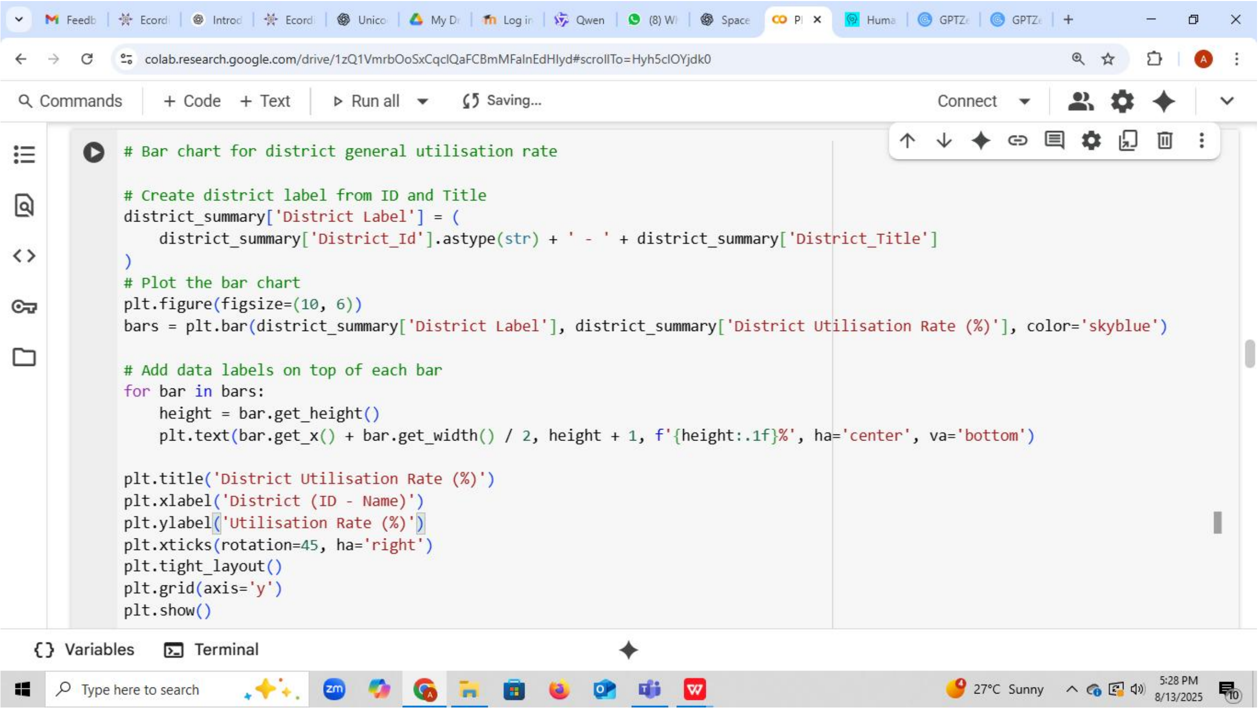


Fig. 24.3: District utilisation rate bar chart_1

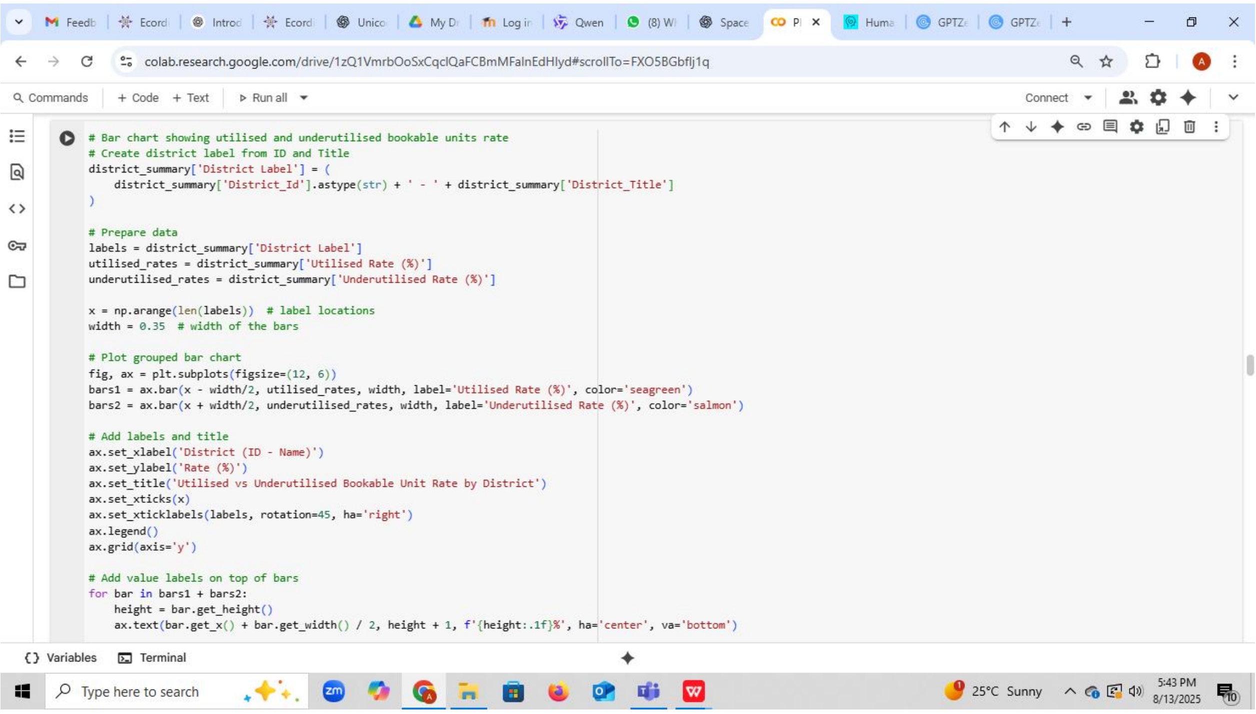


Fig. 25.1: Utilised vs underutilised bookable units by district_1

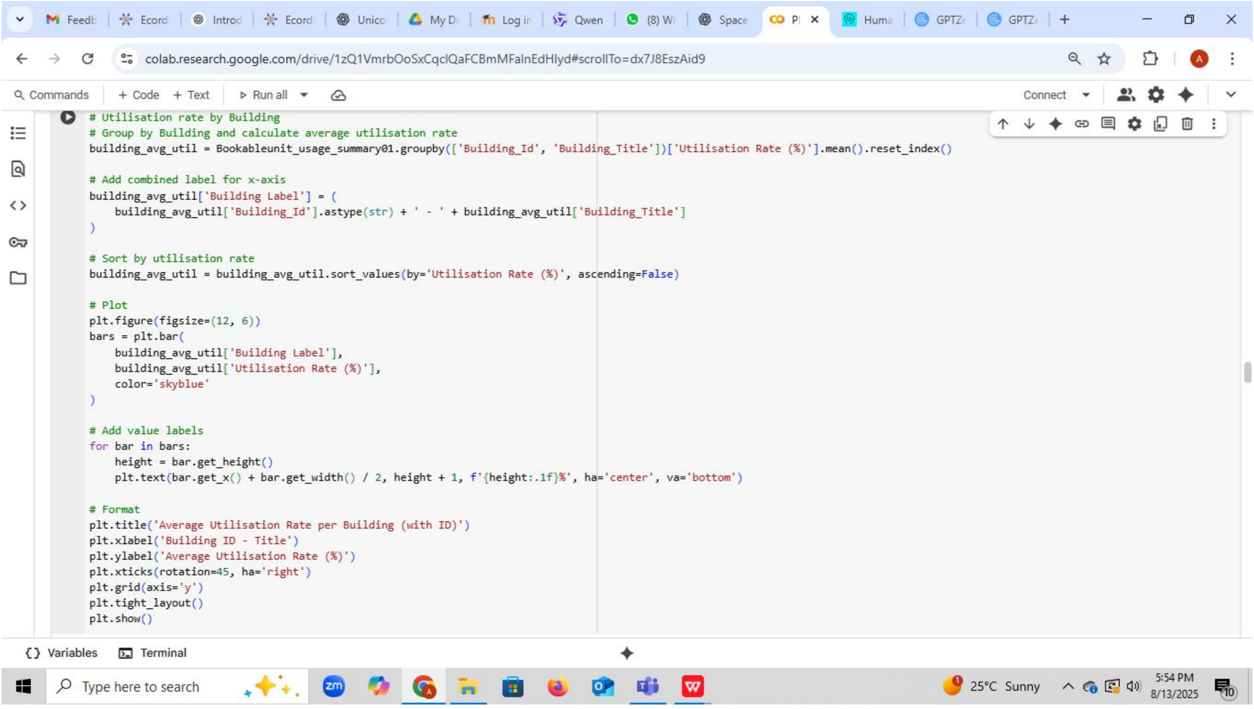


Fig. 26.1: Building utilisation rate_1

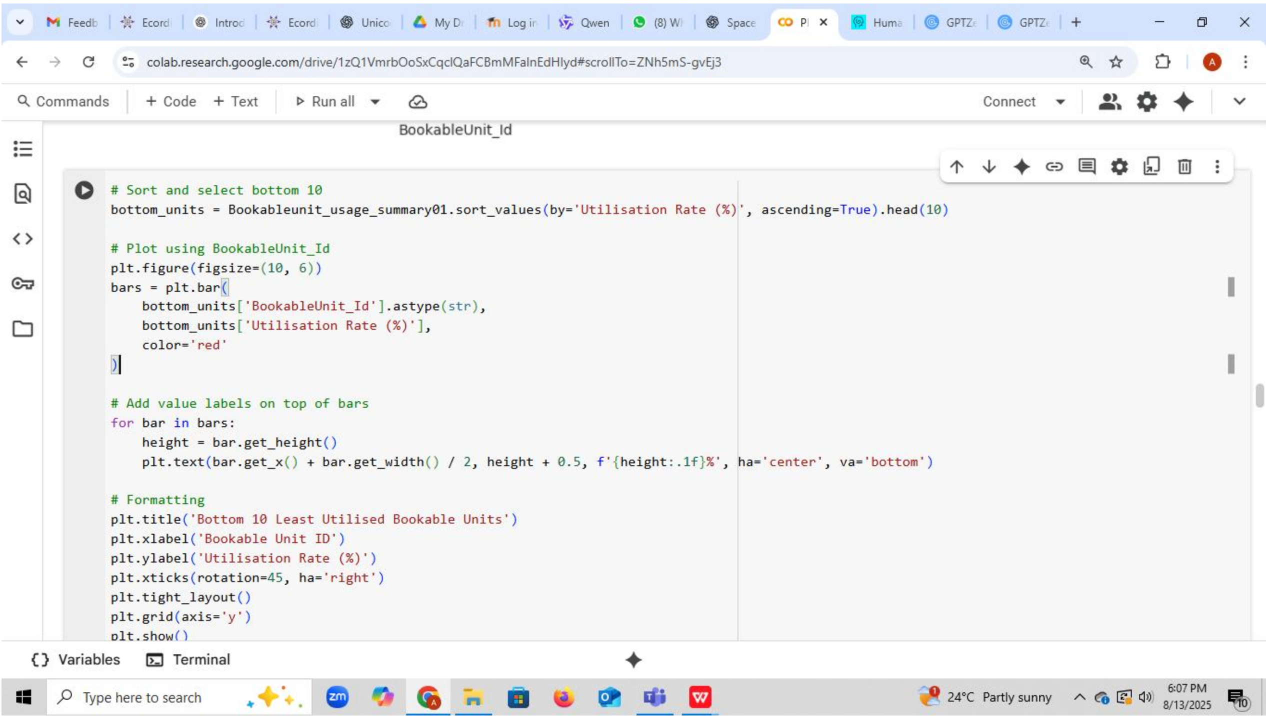


Fig. 28.1: 10 least underutilised bookable units_1

Appendix D: Correlation Analysis

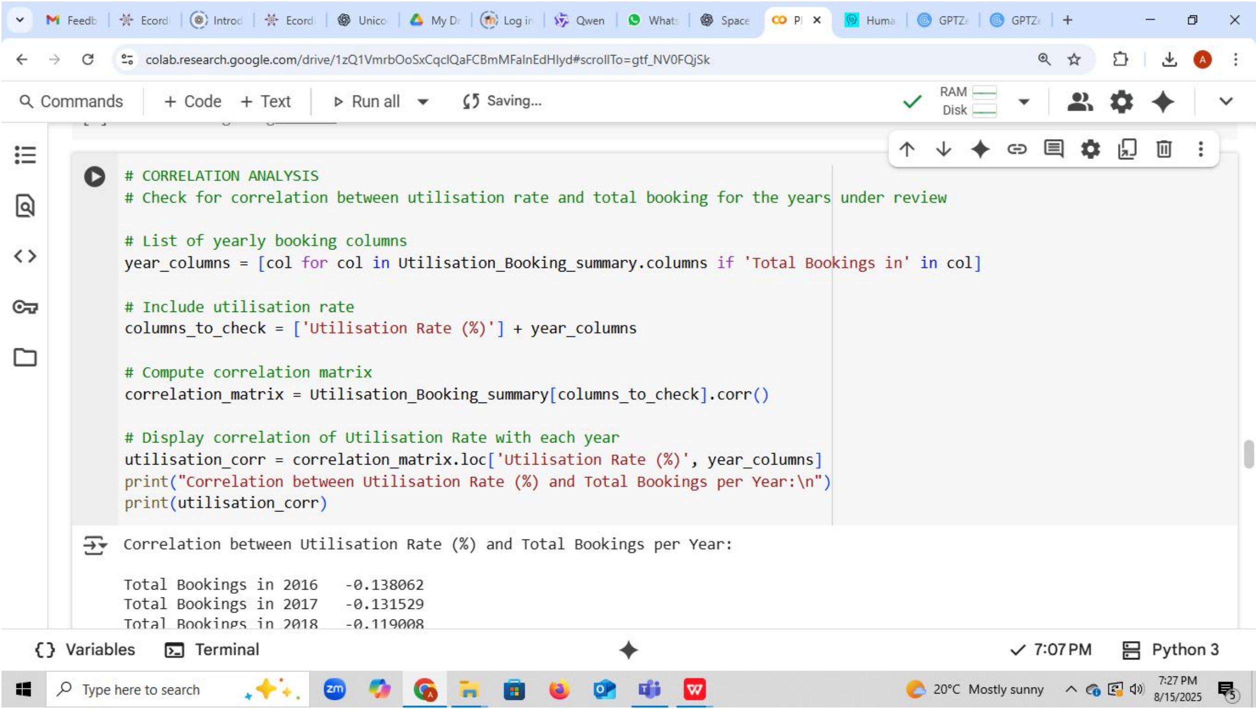


Fig. 29.1: Correlation analysis between utilisation rate and total yearly booking_1

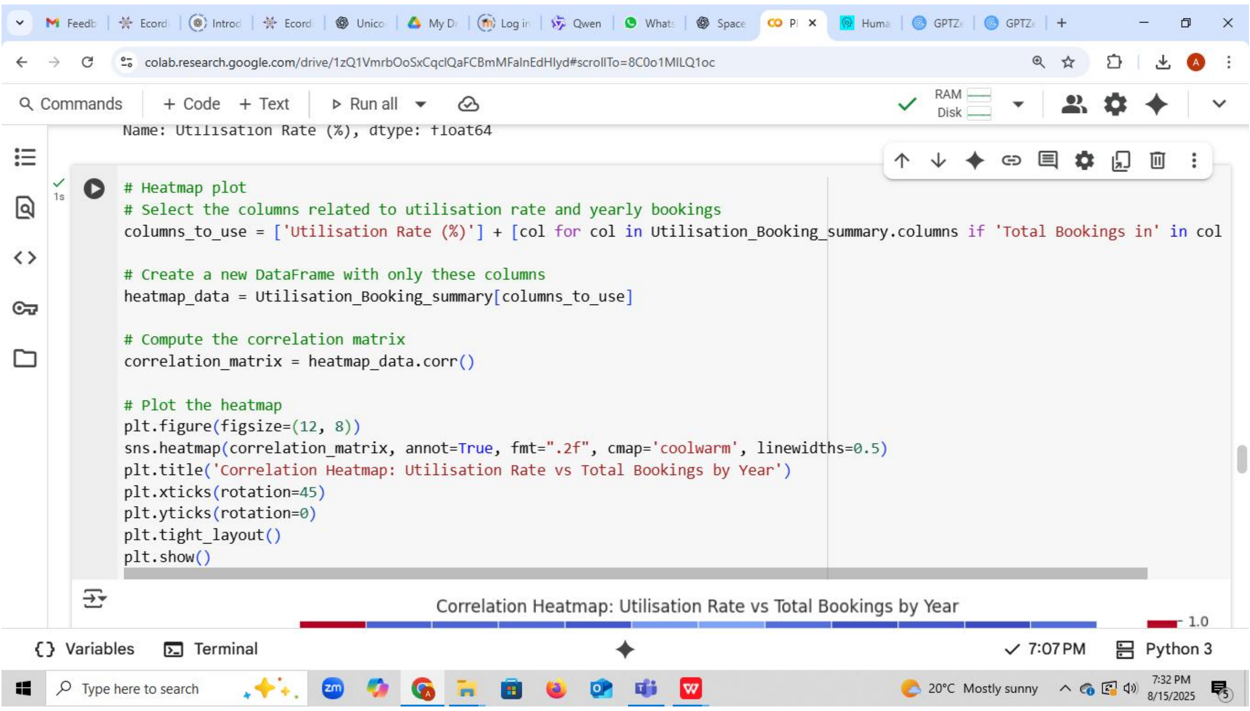


Fig. 30.1: Heatmap plot_1

Appendix E: Summary of Maywoods Booking dataset after evaluation

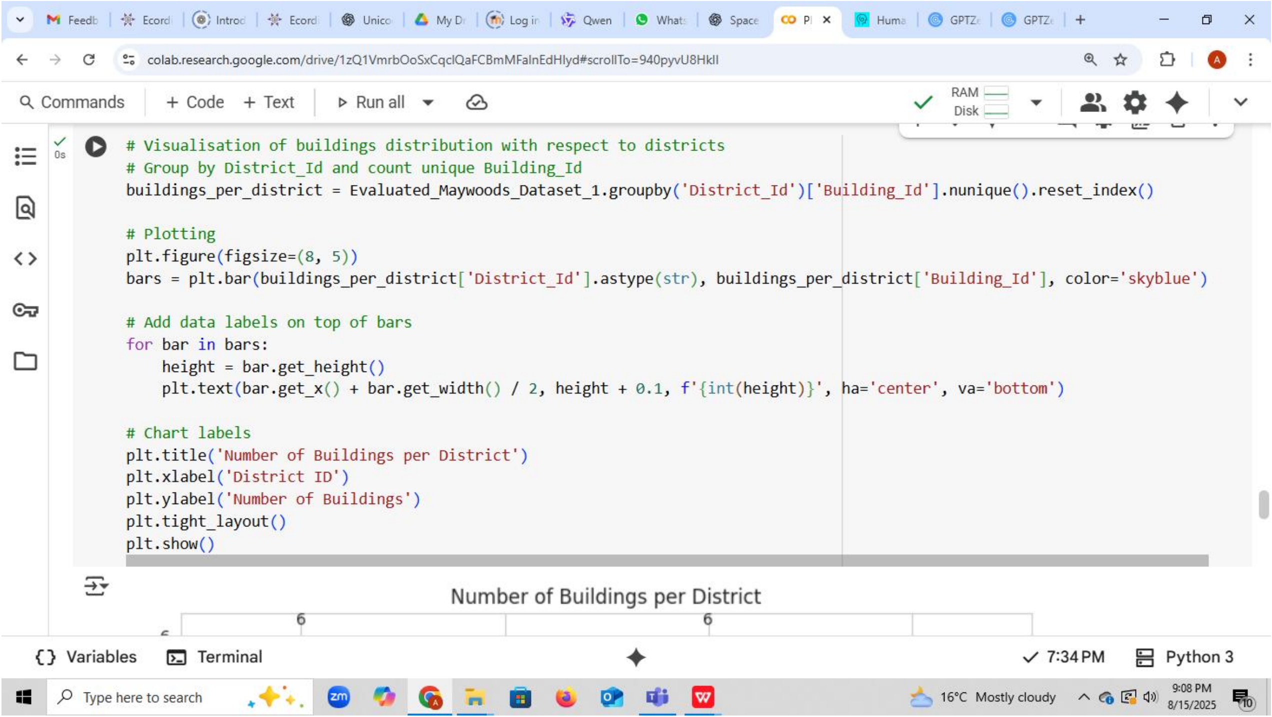


Fig. 34.1: District bar chart after evaluation_1

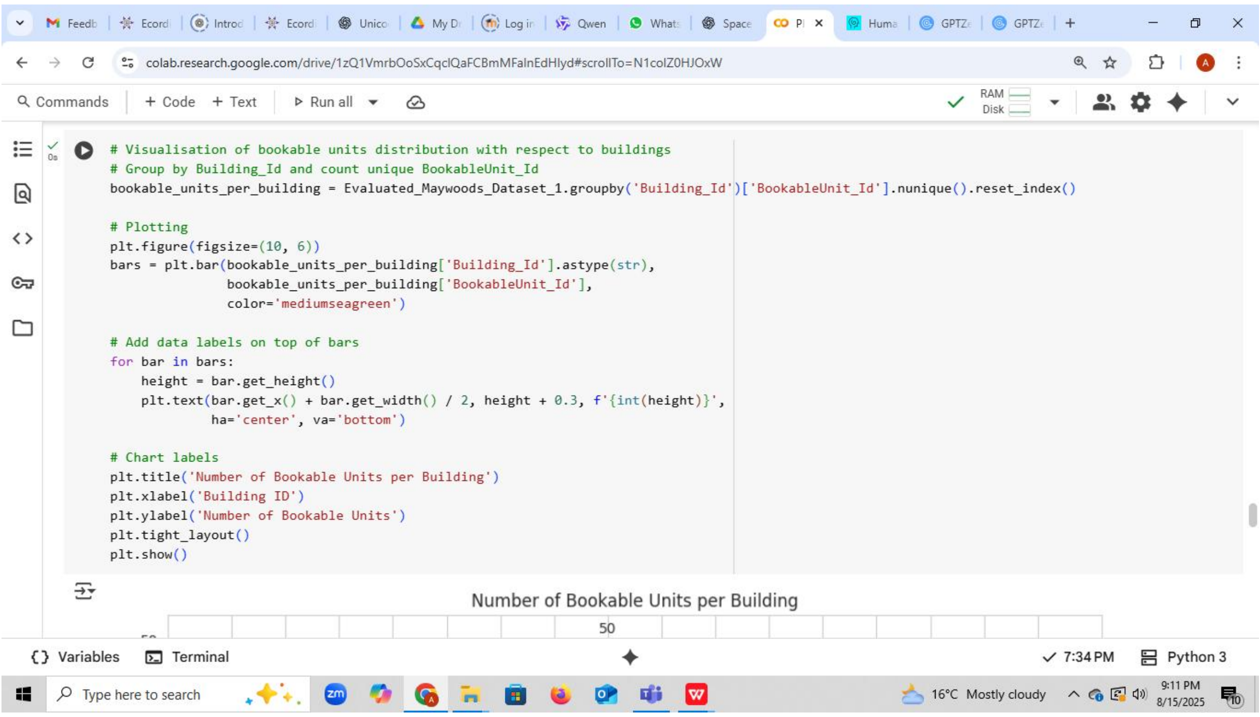


Fig. 35.1: Building bar chart after evaluation_1

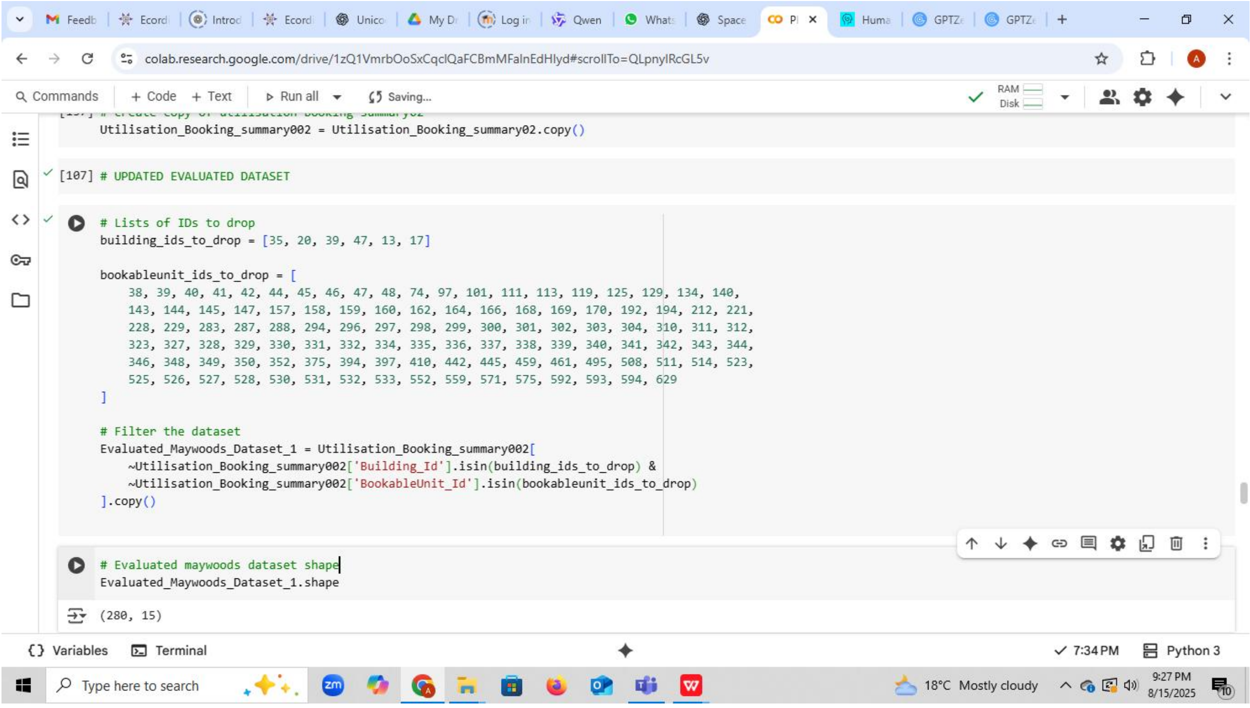


Fig. 36.1: Evaluated maywoods dataset_1

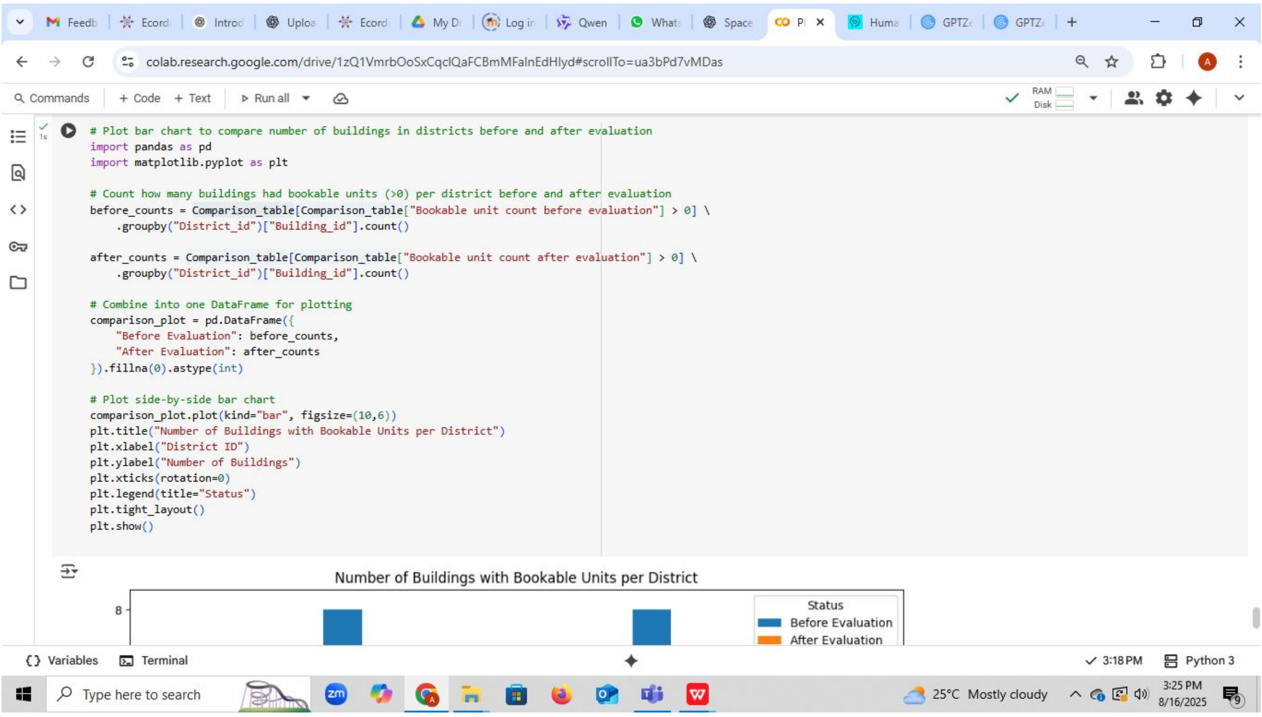


Fig. 37.1: Comparison of districts before and after evaluation_1

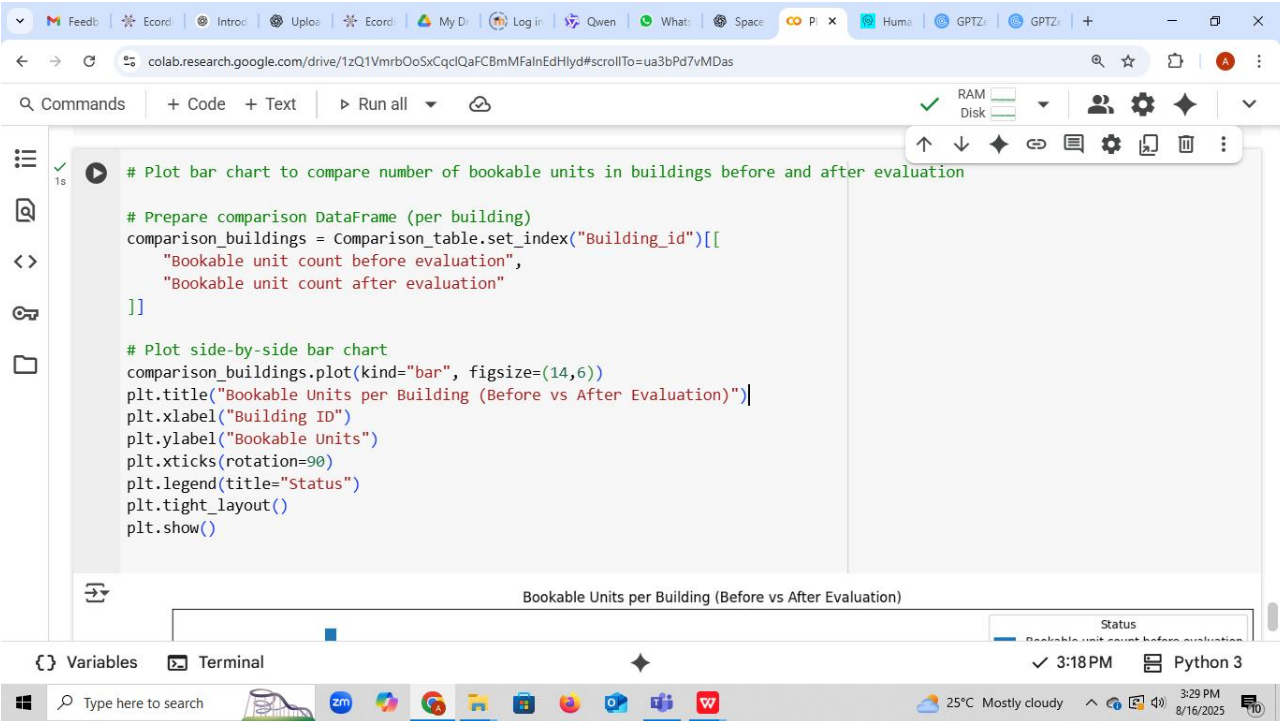


Fig. 38.1: Comparison of buildings before and after evaluation_1

Appendix F: Predictive Modeling

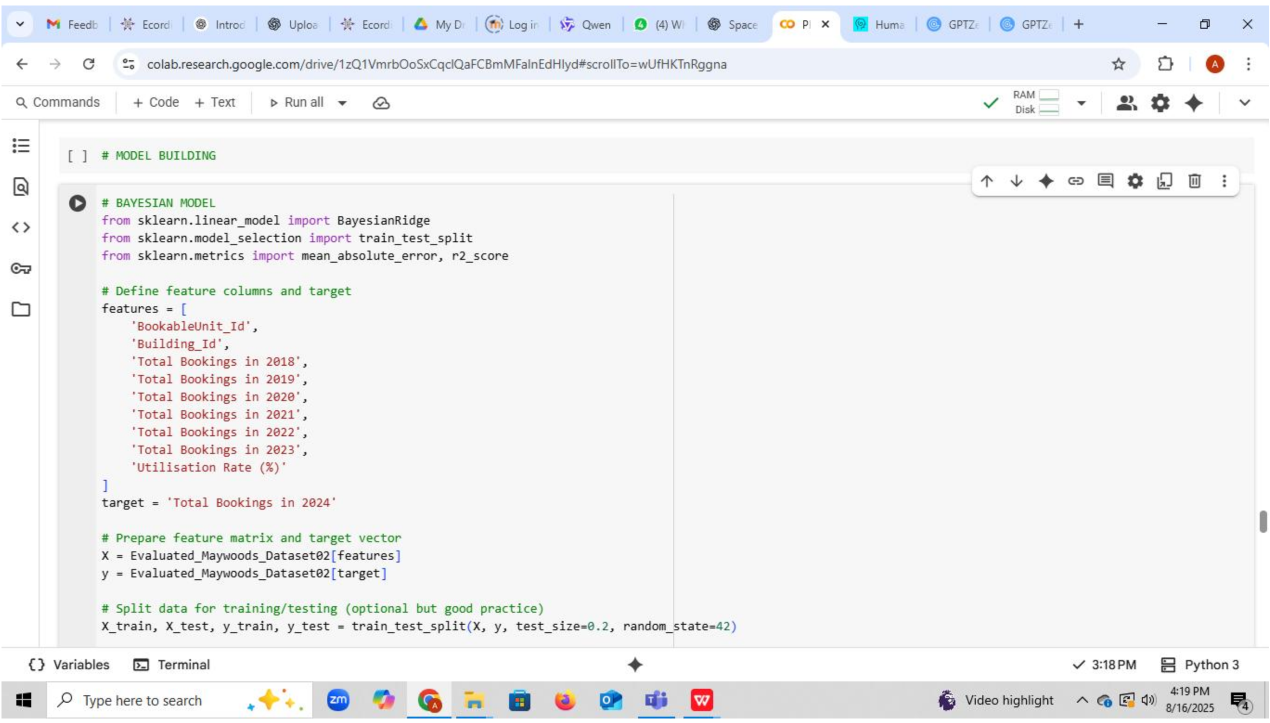


Fig. 41.1: Bayesian ridge regression model training and testing_1

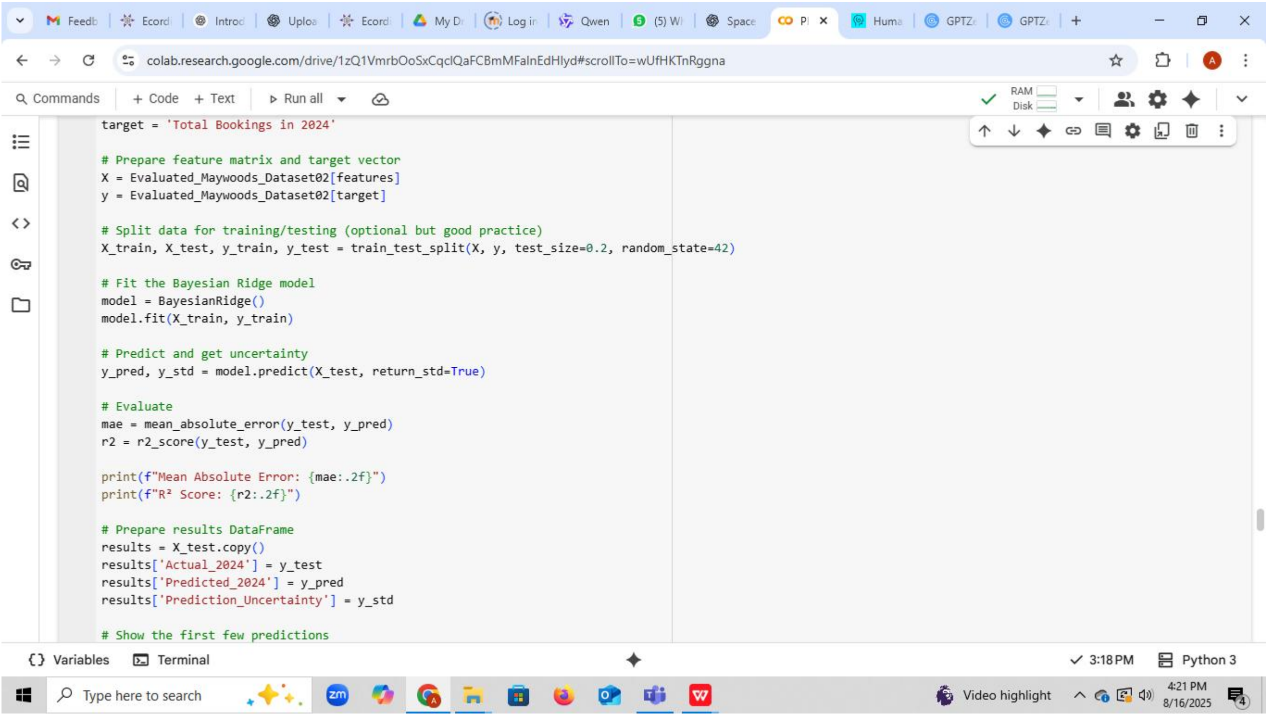


Fig. 41.2: Bayesian ridge regression model training and testing_2

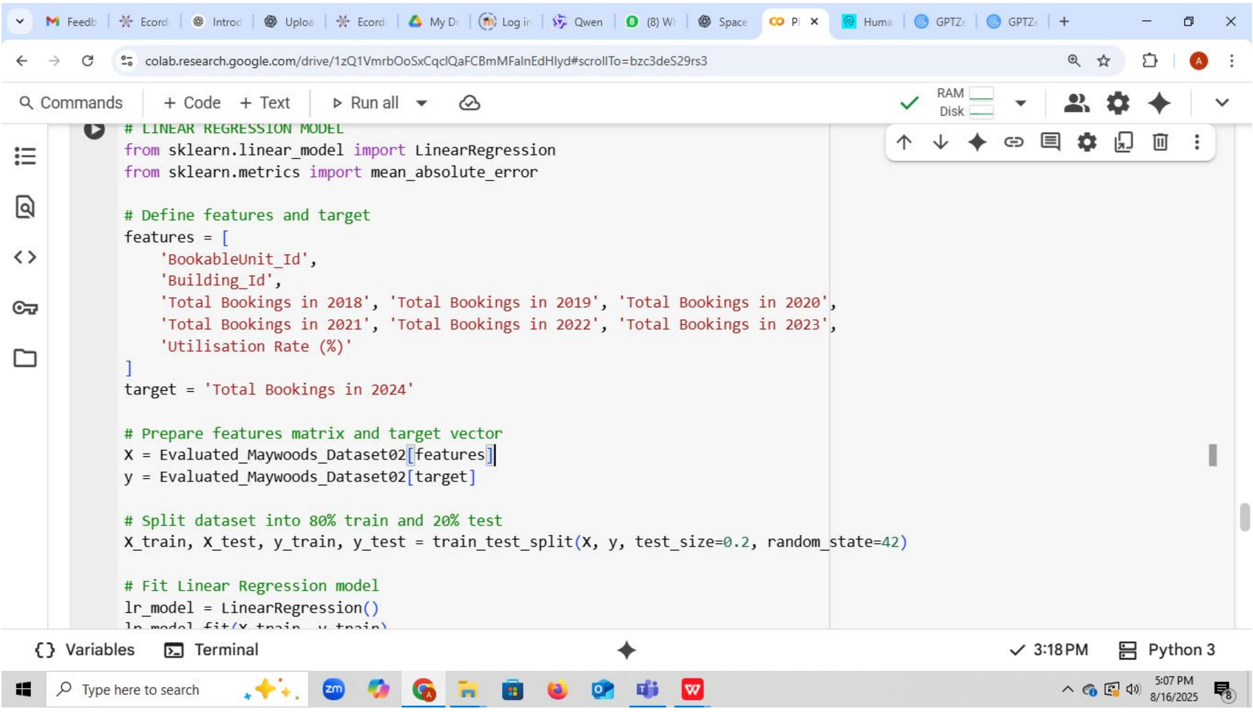


Fig. 42.1: Linear regression model training and testing_1

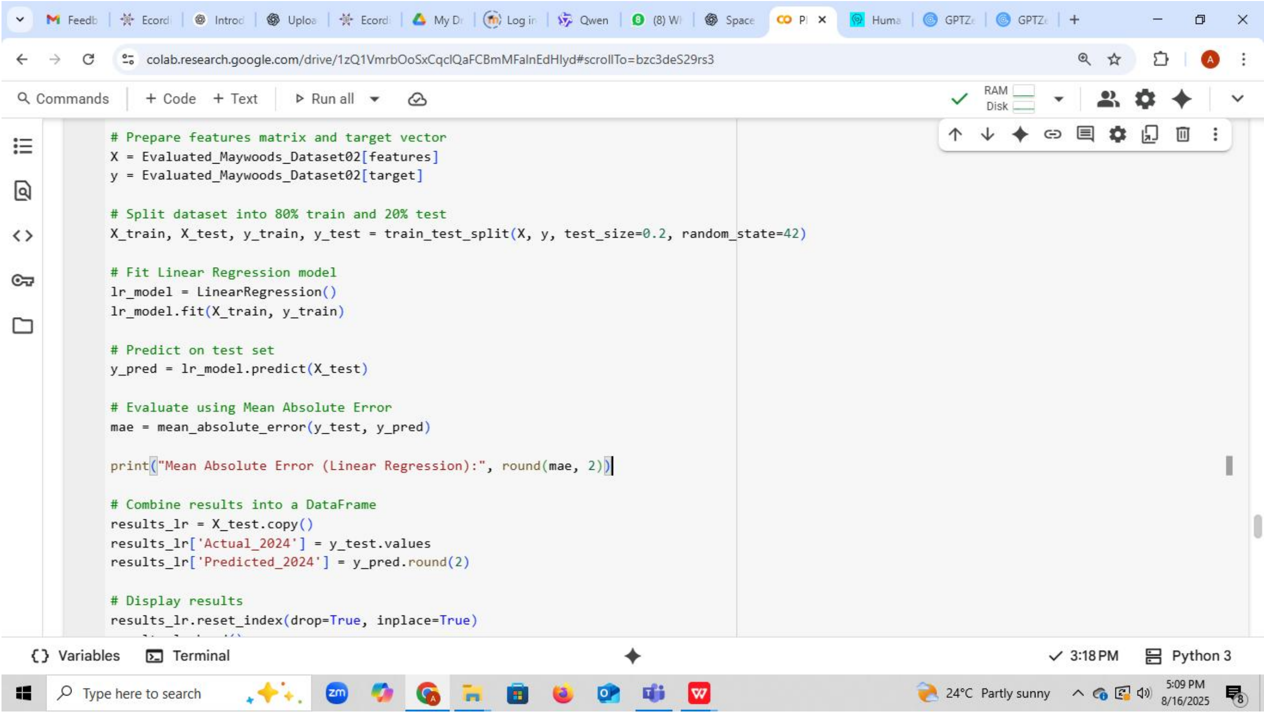


Fig. 42.2: Linear regression model training and testing_2

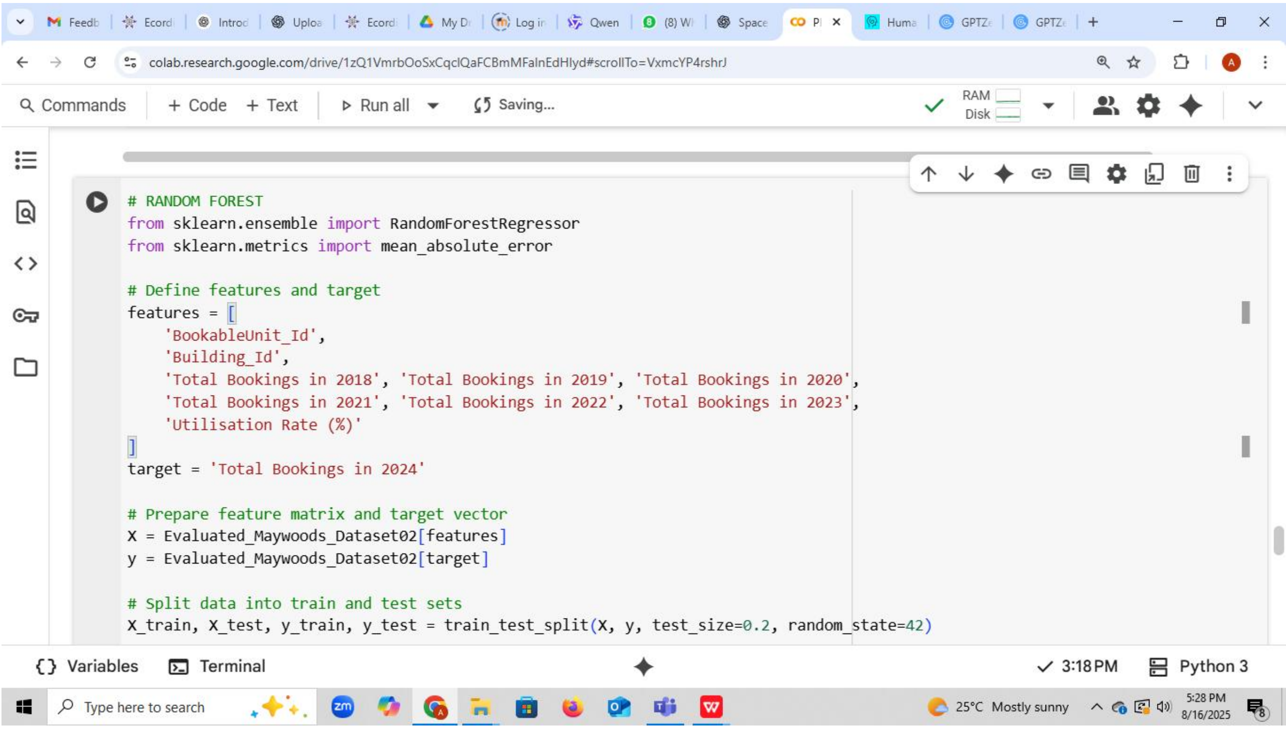


Fig. 43.1: Random forest training and testing_1

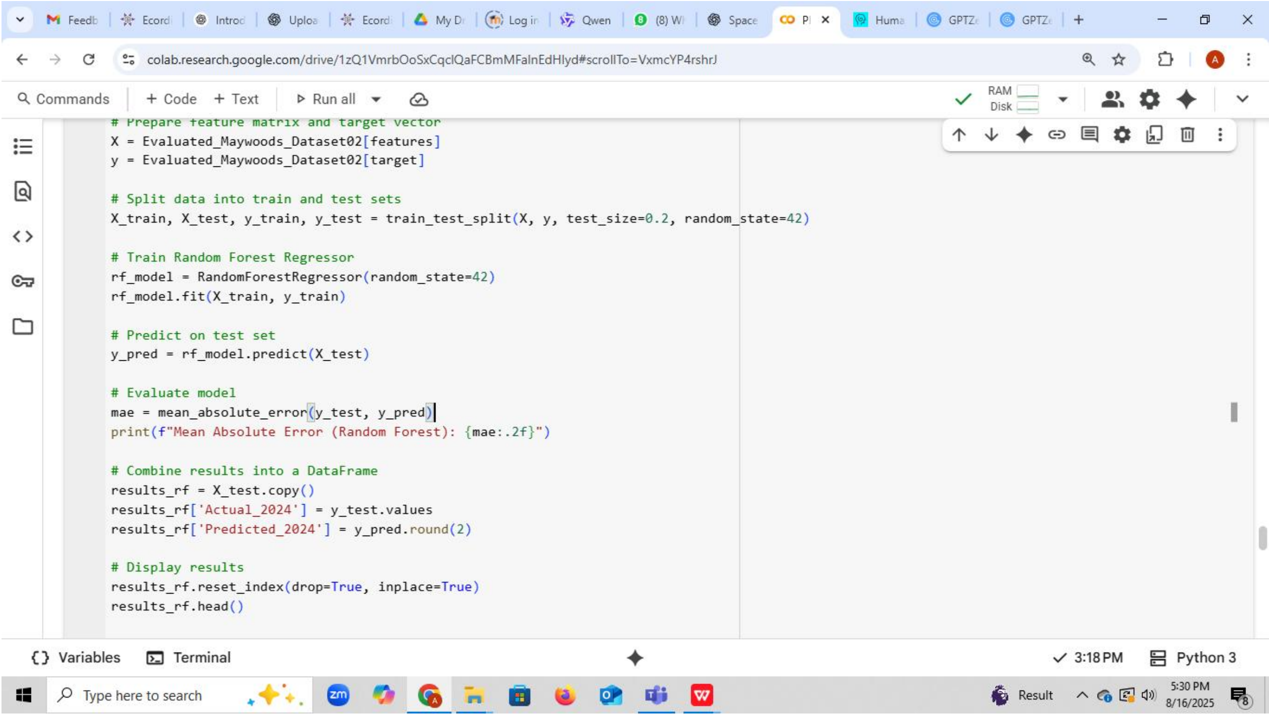


Fig. 43.2: Random forest model training and testing_2

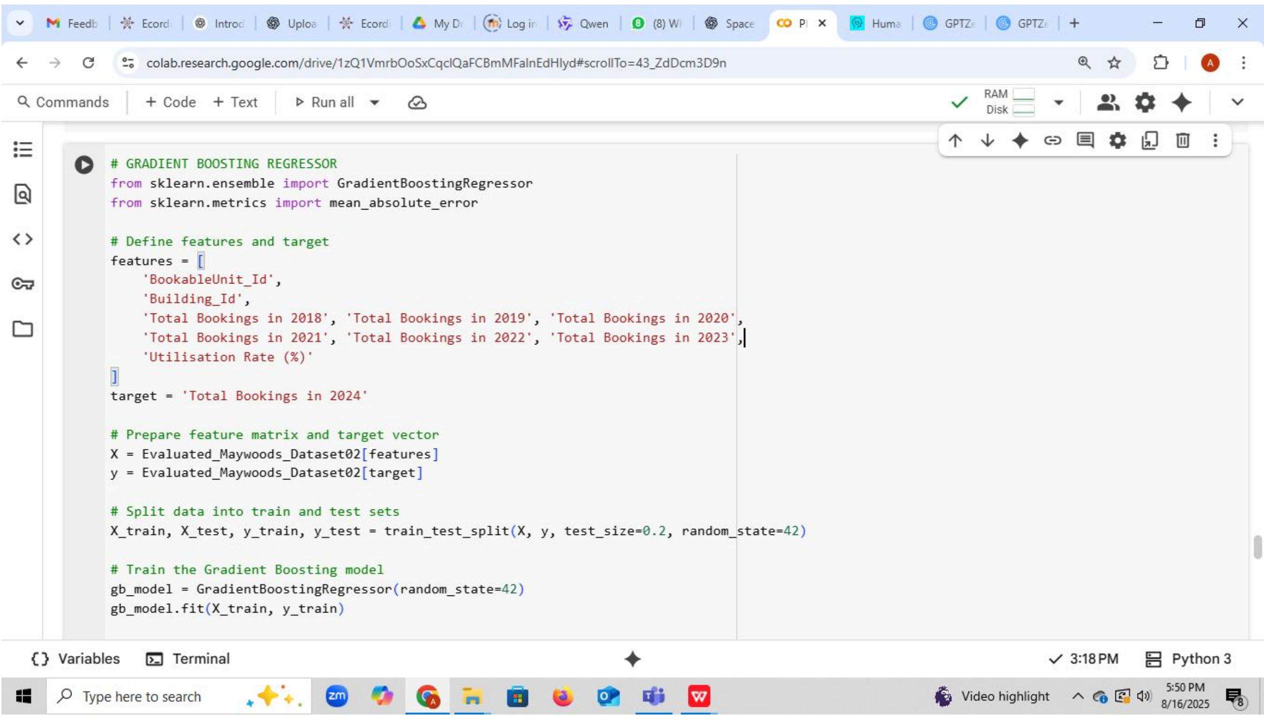


Fig. 44.1: Gradient boosting regressor training and testing_1

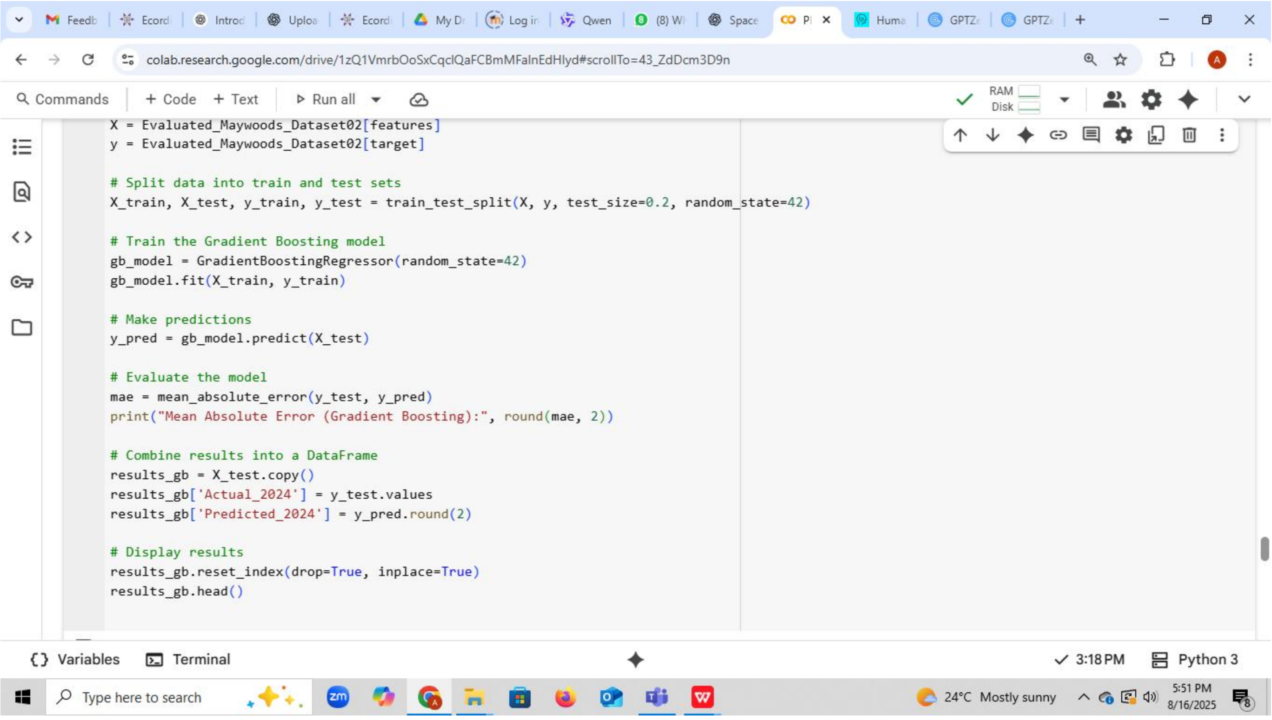


Fig. 44.2: Gradient boosting regressor training and testing_2

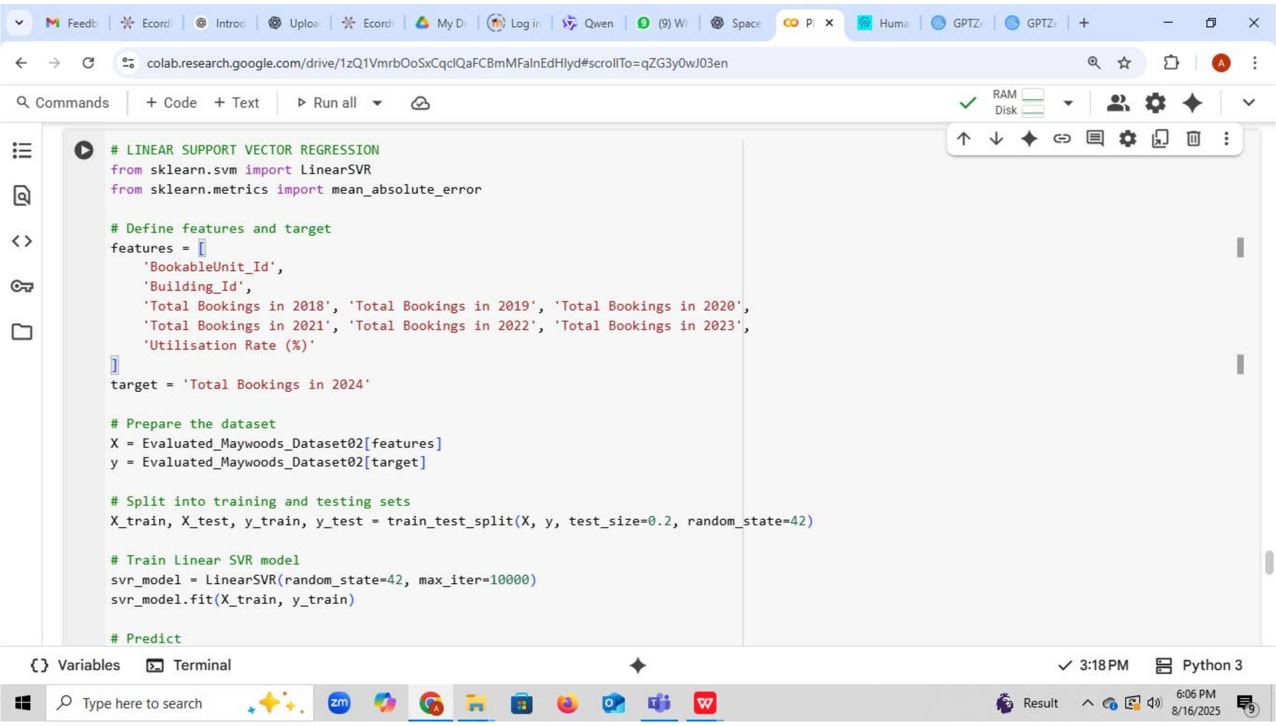


Fig. 45.1: Support vector regression training and testing_1

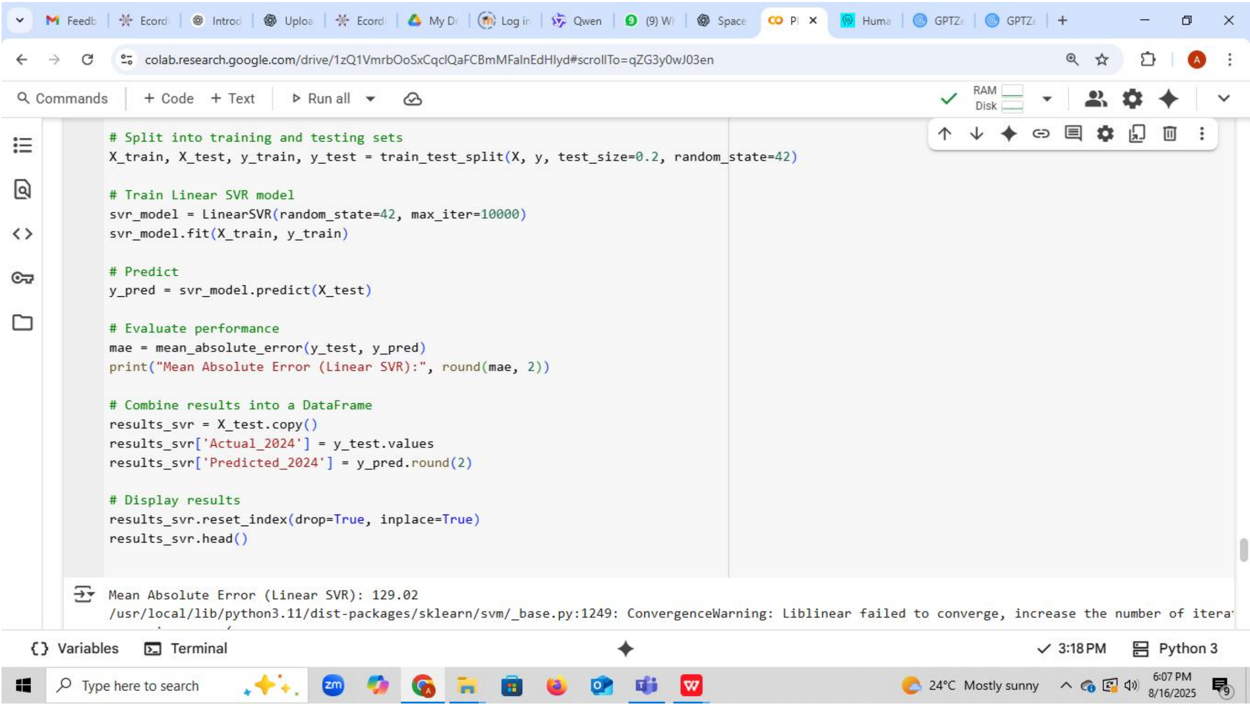


Fig. 45.2: Support vector regression training and testing_2

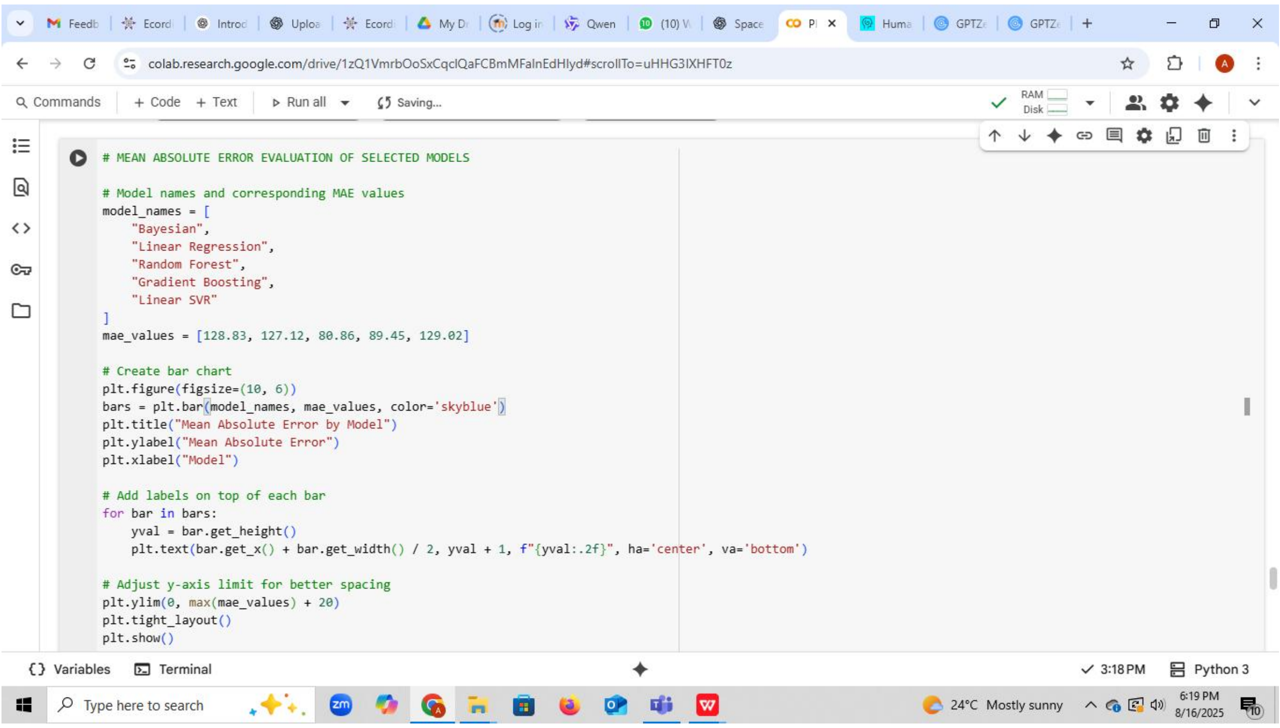


Fig. 46.1: Model evaluation_1

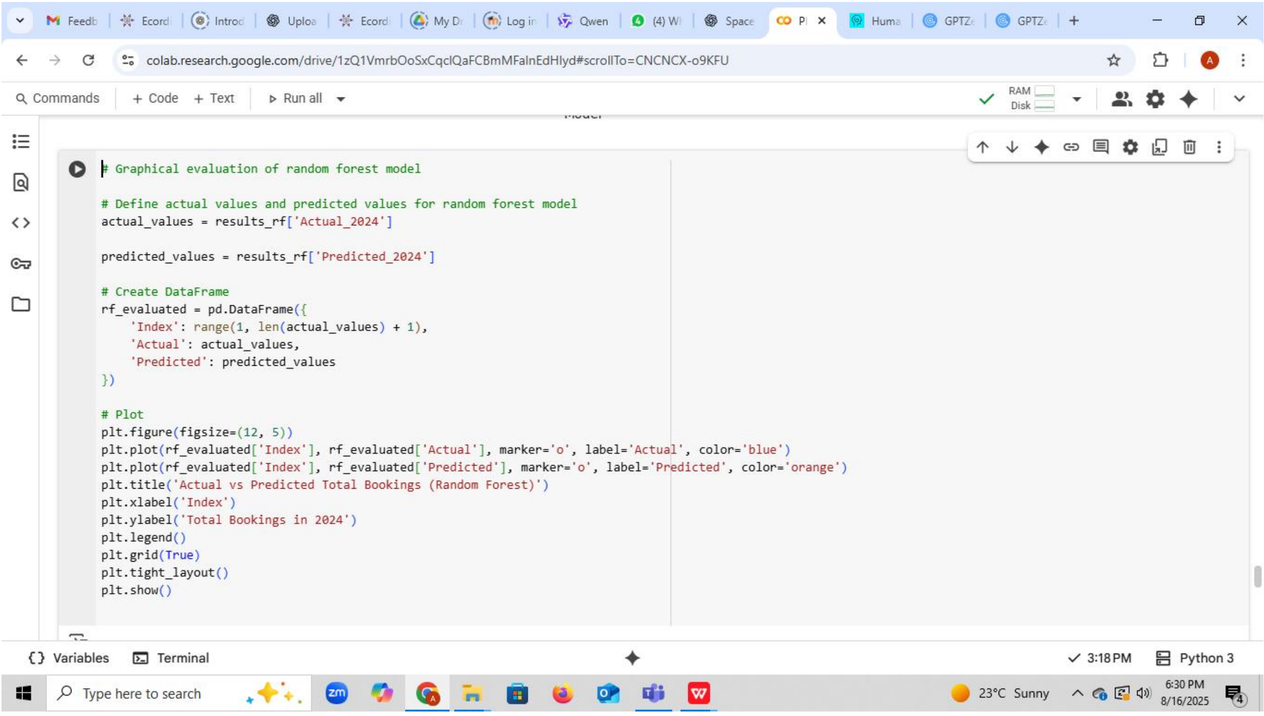


Fig. 47.1: Random forest model evaluation plot_1

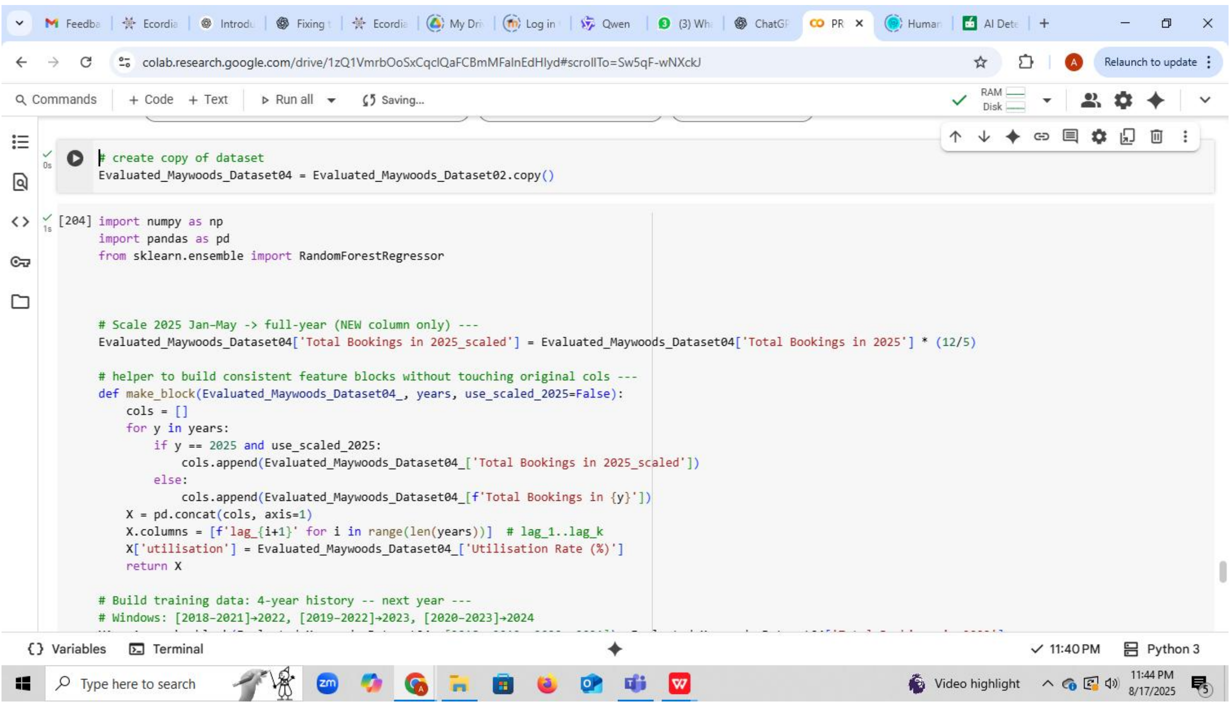


Fig. 48.1: Forecast for total booking 2026 (1)

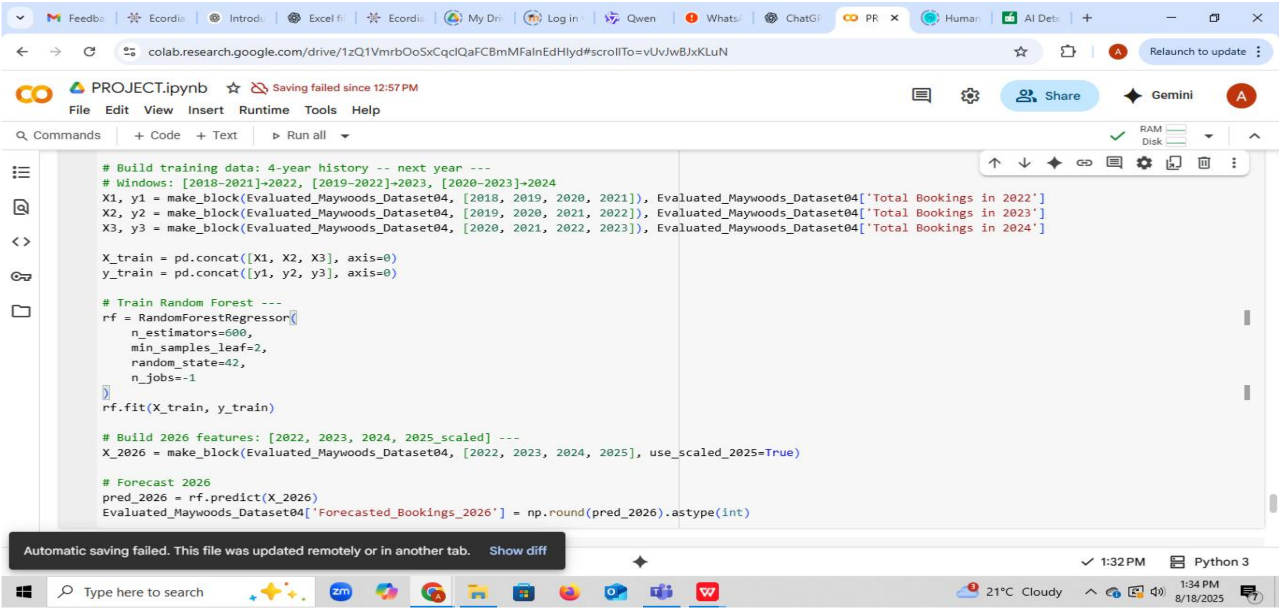


Fig. 48.2: Forecast for total booking 2026 (2)

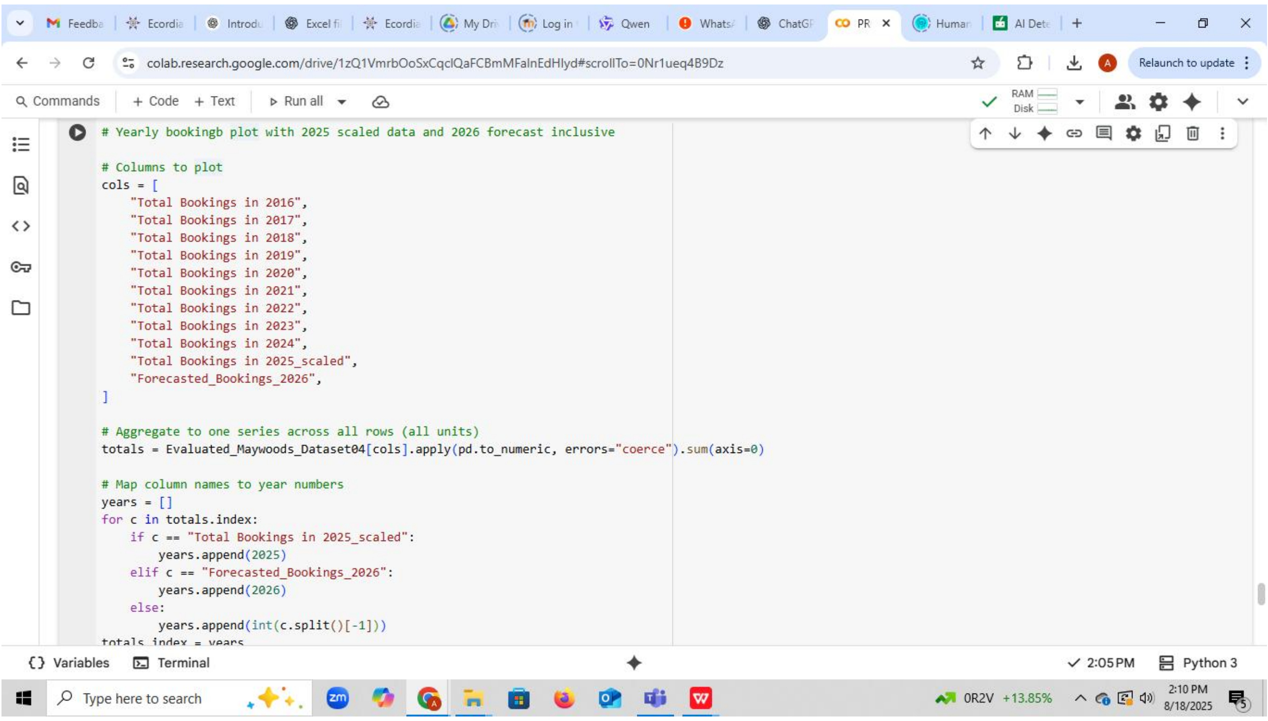
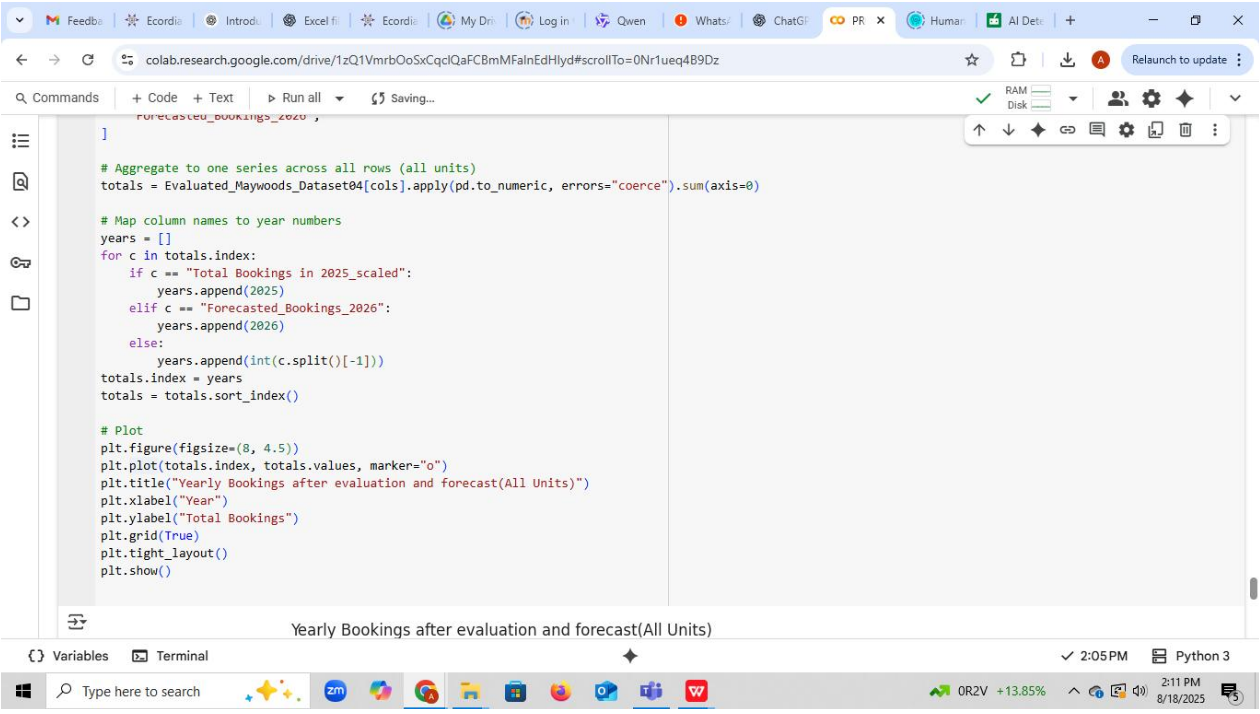


Fig. 49.1: Yearly booking plot with 2026 forecast inclusive (1)



```
    "Forecasted_Bookings_2026",
]

# Aggregate to one series across all rows (all units)
totals = Evaluated_Maywoods_Dataset04[cols].apply(pd.to_numeric, errors="coerce").sum(axis=0)

# Map column names to year numbers
years = []
for c in totals.index:
    if c == "Total Bookings in 2025_scaled":
        years.append(2025)
    elif c == "Forecasted_Bookings_2026":
        years.append(2026)
    else:
        years.append(int(c.split()[-1]))
totals.index = years
totals = totals.sort_index()

# Plot
plt.figure(figsize=(8, 4.5))
plt.plot(totals.index, totals.values, marker="o")
plt.title("Yearly Bookings after evaluation and forecast(All Units)")
plt.xlabel("Year")
plt.ylabel("Total Bookings")
plt.grid(True)
plt.tight_layout()
plt.show()
```

Yearly Bookings after evaluation and forecast(All Units)

Fig. 49.2: Yearly booking plot with 2026 forecast inclusive (2)