WARWICK BUSINESS SCHOOL
THE UNIVERSITY OF WARWICK

Student Number	5527700
Module Code	IB2D40
Module Title	Programming for Business Application
Submission Deadline	06-Jan-2025 12:01:00 PM
Date Submitted	05-Jan-2025 11:19:13 AM
Word Count	2216
Number of Pages	9
Question Attempted	The PDF report answers the assignment question 2 from the assignment instruction. The 6 sections of the report give answers of the 5 subset questions (a), (b), (c), (d) and (e) from the question 2 and a reference list. The topic of each section has been indicated in bold at the beginning of the section, matching the subset question it answered. The entered word count for the report does not include neither the reference list nor the page numbers at the bottom of the each page, being read from its Word docx version. The front size and formats of the report have strictly complied the WBS word count policy.
Have you used Artificial Intelligence (AI) in any part of this assignment?	No
If you have ticked "Yes" above, please briefly outline below which AI tool you have used, and what you have used it for. Please note, you must also reference the use of generative AI correctly within your assessment, in line with the guidance provided in your student handbook.	

a. Step Description

After importing Pandas library, during data pre-processing I firstly renamed the columns from column 0 to 6 to "day", "Hour of Arrival", "Road traffic accident", "Assault", "Deliberate self-ham", "Sports injury" and "Not known" respectively within the data frame loaded in the form of .xlsx. Afterwards I deleted the redundant rows containing no data. While cleaning the data, I checked unique values for each categorical column. There were missing values in column "Sports injury" and "Deliberate self-harm". To handle the missing values, I converted all values in the two columns to numeric, converted the non-numeric to missing values, and replaced the missing values in each column by their averages respectively. I removed all duplicated rows in the data frame and only one row remained.

To analyze the first insight, the hour had the highest total attendances in the week, *hourly total attendances* column was created, summing attendances for 5 reasons in each hour of each day. Then, I grouped and summed the number of hourly total attendances by hours regardless of days, storing the output as a new variable *weekly_total_attendances.* Attach ".idxmax()" to *weekly_total_attendances* to find the hour had the most attendances.

For analyzing the second insight, the highest proportion of the attendance reason for everyday and the lowest, I created a data frame *daily_total_attendances* involved new column *hourly total attendances* and summarized the number of attendances on daily basis, concluding the 24/7 hourly attendances to 7 days. The frame would be deployed with double *for* loops iterating a list of 5 attendance reasons and a list of 7 days.

The loop iterated the reason list inside the loop iterating the day list to calculate the percentage of each attendance reason in each day. Under the "reason" *for* loop inside the "day" *for* loop, I calculated the proportion of the number of each attendance reason to the total attendances in each day and stored the proportions to a dictionary called *proportion*. Under the same "reason" loop, the proportions of reasons in the same day appended to a variable called *Proportion_of_the_day*. After finishing the interior loop, "max()" and "min()" functions to *Proportion_of_the_day* under the "day" loop will find out the attendance reasons with the highest and lowest proportion on same day. The corresponding day's name will be matched by searching the *proportion* dictionary.

Proportion_of_the_day variable was an empty list inside the "day" loop and before the "reason" loop. The variable has same indentation length as the "reason" loop. This position emptied the list variable each time at the beginning of instructions to each iterator day.

The third key insight is the midnight effect on attendance reasons, examining how hourly attendances vary during the time joint from 11 p.m. to 1 a.m. the next day. The insight was demonstrated via the attendances pattern during the time joint on average and individual time joints during the week. I transformed the number of hourly attendances by reasons and calculated the total attendances characterized by hours, resulting in *hourly_attendance_summary* data frame. Then I attached "unique()" to output categorical values within column *Hour of Arrival* and converted to a list by "tolist()". Moreover, I located the 23:00 and 1:00 in the hour list via indexes and saved them to facilitate the analysis.

In terms of the average pattern I located the data sets in *hourly_attendance_summary* by ".loc[]" function and using targeted hours as input to find the attendances at 23:00 and 1:00 and divide the numbers by 7 to solve the attendances on weekly average. Calculating the difference between the average attendances at the 2 hours showed the average pattern of the midnight effect on attendances.

Alternatively, addressing midnight effects during individual time joints paired contiguous days in a week, for instance the attendance changes between 23:00 Tuesday and 1:00 Wednesday. For the accuracy of data extraction, deploying a *for* loop iterating "day" through 7 days conditioned the assignment to variables "next_day" and "day" for each loop run, a conditioned dictionary maps each day to its next day such as the next day of Sunday is Monday. "df['day']" and "df['Hour of Arrival']" looked up the data set satisfying conditions on day that loop iterated and conditions on arriving hours, finding the data at the time joint. ['hourly total attendances'] following the lookup specified the column data in the targeted data set. ".values[0]" extracted the targeted value from the specified column by converting the format Series into a NumPy array. The number at 1 a.m. the next day minus the number at 11 p.m. on the last day showed the midnight effect during the specific time joint. And the *for* loop iterating the 7 days calculated how the attendances changed individually during every midnight of the week.

b. Insights

The objective of the NHS decision-making is to maximize the efficiency of resource allocation, which allocates the correct quantity of needed resources to meet patients' medical demands, maximizing the utility from resources. The resources in this report refer clinical services, people such as doctors and nurses, clinical support services like pharmacy, logistics such as procurement and facilities, and finance (NHS improvement, 2017).

Insight 1 implied demands for medical services peaked at 9 a.m. on average. The NHS should allocate most of its resources to that period every day to handle the surge in Accident and Emergency (A&E) admissions, improving the allocation efficiency of resources. The NHS could increase the capacity for clinical services and clinical support by increasing the hospital beds and pathology apparatus available at 9 a.m. to conduct medical treatments. The NHS could also arrange the schedules of medical professionals to be mostly available at that time to ensure there is sufficient labour to provide care. In logistics area, the NHS could strengthen the maximum capacity of hospital data systems to prevent collapsing at that busy time because of the highest number of attendances and the servers suffering the most requests, the decision guarantees hospitals could function normally and provide medical services to patients in time during busy time. Simultaneously, the NHS could lower the resource allocations during periods with fewer attendances, reducing the expenditure and reallocating the saved resources to busier periods to boost the efficiency of resource allocation.

Insight 2 helped specify resource allocations to certain categories to improve efficiency via the most frequent reason and the least frequent one. The most common reason was "Not known" for weekdays and "Sports injury" for weekends, the least common reason was

“Deliberate self-harm” during the whole week. Apart from “Not known” was unidentifiable for medical demands, the NHS could allocate more resources to hospital departments related to caring sports injuries such as surgery department on weekends due to the insight, for example scheduling more surgeons and renting more wheelchairs placed at hospitals. Also, the NHS could arrange a smaller number of resources to tackle patients with deliberate self-harm issues, scheduling fewer psychiatrists and buying fewer medicines targeting psychosis proportionally to the attendances. These arrangements regarding reasons for A&E attendances make sure the needed resources are allocated to most patients at the correct time and prevent the waste of resources, improving the allocation efficiency of resources.

Insight 3 showed midnight effects on attendances, and the NHS could reallocate labour and facilities during the midnight to improve efficiency of resources by the insight. Data analysis suggested all the 7 time joints had a decline in attendances from both average and individual perspectives. The NHS should schedule fewer staff and allocate fewer support facilities during the midnight in response to the diminishing attendances. According to the NHS Employer (2022), pay rates for staff working at midnight are at least 30% higher than normal working time across all pay bands, incurring a large expenditure on the NHS budget. This decision therefore cuts the expenditure on redundant labour and accompanied hospital facilities such as electricity and water supply. Consequently, the NHS will maximize the allocation efficiency of resources during midnight and reduce the NHS financial pressure.

c. Visualizations

Fig.1 Peak hour of attendances and Midnight effects

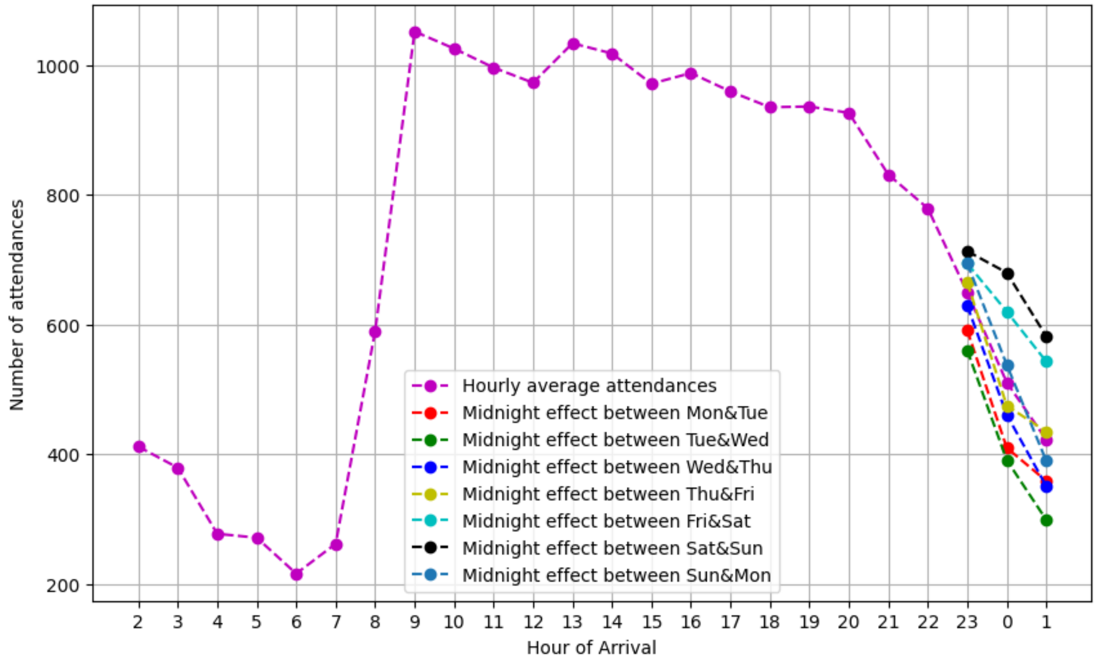
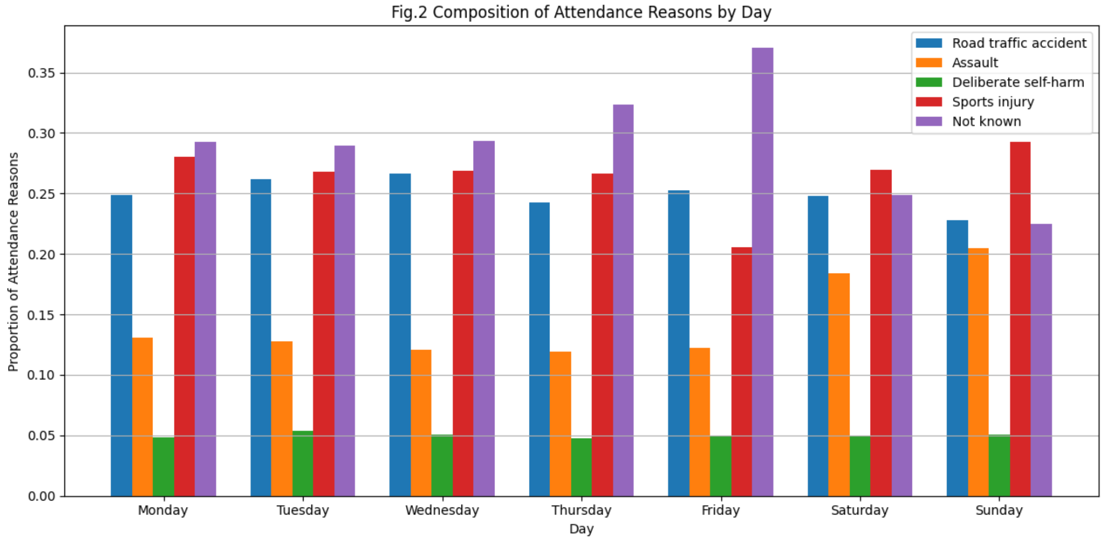


Fig.2 Composition of Attendance Reasons by Day



d. Discussion and Proposal

Regarding the issue in insight 2, “Not Known” reason category has occupied the highest percentages during all weekdays while the NHS cannot know how to improve its planning on A&E resources from this outcome. This issue potentially contributed to the allocation inefficiency of resources and the failure to meet the largest patients’ demand. An investigation on reason category is therefore needed to collect the detailed attendance reasons within it and add new categories to deconstruct “Not Known” such as “Fever”, “Collapse” etc. Alternatively, the NHS could change the grouping methodology to group the attendances by hospital departments instead of grouping for attendance reasons, patients therefore can be sent to departments they need in order to be diagnosed and remedied more rapidly.

With respect to the issues in insight 1 and insight 3, it is not clear when the patients were receiving medical treatments at hospitals because the *Hour of Arrival* column in the data frame records hours of a given day when the number of admissions was reported, not the hours when the patients were receiving treatments. This disparity causes the insight 1 may not reflect peak hour correctly and the insight 3 may not reflect correctly the change in attendances during midnight effect since it could take time for hospitals to upload the numbers to the repository and the admissions may have happened before the records in *Hour of Arrival*. These issues could invalidate the analysis results and NHS allocation of resources, being not time sensitive. Hence a program could be designed to collect patients' admission information to repository immediately through hospital systems when they are being treated, including time of receiving treatment and attendance reasons.

Whereas it is difficult to have reason categories representing all attendances without missing. Research about expressions on hospital admission reasons displayed the difficulties on grouping A&E attendances by reasons, over half of the patients could not state their reasons and reported different ones from their records (Berger, Dembitzer and Beach, 2013). It is therefore ineffective to group attendances by self-reported reasons. Also, grouping the reasons from patients' records done by doctors is hard to determine how specific the categories should be since a problem can split into many subdivisions. For instance, sport injury in surgery comprises thoracic surgery, orthopedics and neurosurgery etc. Hence it is challenging for grouping for either attendance reasons or by hospital departments, increasing the unreliability and uncertainty in data sources.

While maintenance of the software development cycle, my program can separate to features serving data cleaning, analysis and visualization respectively through distributed version controls. A combination of incremental development and agile methodology would be the continuous pattern of increasing functionalities and quality of the program.

The incremental development allows adding functions to clean such as cleaning outliers and text data, functions to analyze such as hypothesis testing, and functions to visualize data such as producing 3D heat maps and spatial graph of attendances, to produce more insights. The added insights can be efficiencies of hospitals in London that measured by the number of attended households and visualized by a spatial heat map, the NHS hence could build more hospitals and allocate more NHS resources to the most demanding regions to decrease medical pressures and allocate less resources to regions where medical demands are low to cut costs.

The agile methodology enables the insights to be more accurate, reliable and time-sensitive in practice by adapting hospitals' feedback over short iterations. For instance, the weekly-iterated feedback on the quantity of resources used for each attendance reason during each period would help the program calibrate its data pre-processing and analysis. If the resources used in peak hours and attendance reasons were not the most, there could be mistakes from analysis or the hospitals' feedback. The program could develop further to use diverse indicators such as patient-to-staff ratios, and time-consistency of peak hours and reasons to justify the validity and reliability of the program over a large time horizon, improving the program quality. The heavy customer involvement ensures the program developers understand the NHS requirements and adapt to its changing needs of diverse attendance reasons like influenza, pandemics etc.

Meanwhile, the distributed version control system saved the program on every developer computer. It allows developers to update codes simultaneously on server-side without interrupting the program on client-side, crucial for making time-sensitive decisions. Also, while the server computer losses due to natural disasters, each computer in the hospital could still use the program to make decisions independently with the decentralized system.

e. Reflection

I searched online resources and experimented with them to overcome coding challenges instead of using AI. The challenges included summaries of data sets by days and hours, targeting data sets through requirements on columns, matching days to correct pairs under a *for* loop, plotting lists of hours on x-axis, naming dictionary keys with loop iterators, and running a *for* loop iterating list indexes and values together. To overcome the challenges, I searched "relevant key words, Python" on browsers to find solutions since browsers could give me a selection of relevant answers from diverse resources such as blogs, videos and communities. Then I would attempt answers from the resource links on my previous codes with IDE to test whether the desired output was returned. If not, I would try changing the details of the code to achieve desired effects and compare codes from other sources.

Reference:

Berger, Z., Dembitzer, A. and Beach, M.C. (2013) 'Reason for Hospital Admission: A Pilot Study Comparing Patient Statements with Chart Reports,' *Narrative Inquiry in Bioethics*, 3(1), pp. 67–79.

NHS Employers (2022) *Unsocial hours payments: Find out about unsocial hours payments, including who they apply to and the rates of pay.* https://www.nhsemployers.org/articles/unsocial-hours-payments.

NHS improvement (2017) *Use of Resources: assessment framework*, *https://webarchive.nationalarchives.gov.uk/ukgwa/20200501111833/https://improvement.nhs.uk/resources/use-resources-assessment-framework/*. NHS improvement. https://www.england.nhs.uk/wp-content/uploads/2020/08/Use_of_Resources_assessment_framework_final.pdf (Accessed: December 22, 2024).