University Health Services (UHS) at UC Berkeley

Consultant: Hin Chi Kwok
Contact: hin-chi.kwok@berkeley.edu

Table of contents

01

Problem Statement

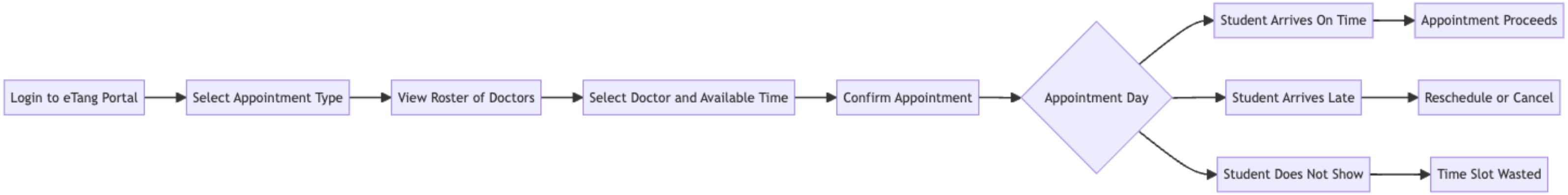
Current healthcare system
Problems and objectives

Healthcare System

UHS Vision:
To be a campus that actively cultivates better health and well-being — a place that can actually make you healthier!

UHS Mission:
To ensure health for all through expert care, leadership, and discovery.

- Current online appointment scheduling system



- Problems and Objectives: reduce the waiting time to improve the operational efficiency and service quality.

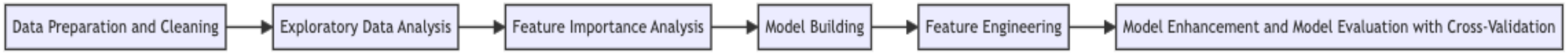
02

Analysis and Methodology

Data Preparation
EDA
Feature Importance Analysis
Model Building

Analysis and Methodology

- Data Preparation and Cleaning
- Exploratory Data Analysis (EDA)
- Feature Importance Analysis
- Model Building
  - Feature Engineering
  - Model Enhancement and Model Evaluation with Cross-Validation



Data Preparation and Cleaning

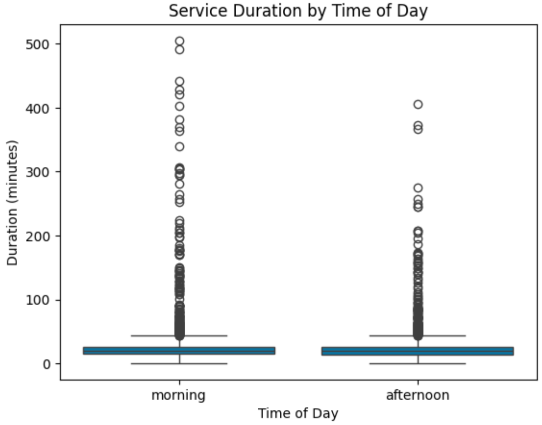
- University Health Services (UHS) statistics originated from fenghaolin/HanguData (github.com) [1, 2]
- Attributes Include:
  - ID, Session, Month, DayOfWeek
  - WorkingDay, AM_PM, Visit.No, Gender
  - Medical Conditions (M.Cancer, S.Cancer)
  - StartTime, PayTime, Address, ServTime
- Key Insights
  - Temporal Data: Data spans across different months and weekdays.
  - Demographic and Health Data: Includes gender and medical condition indicators.
  - Geographical Data: Locations categorized as in-city and out-of-city.
  - Operational Metrics: Service times and session numbers indicate usage patterns.

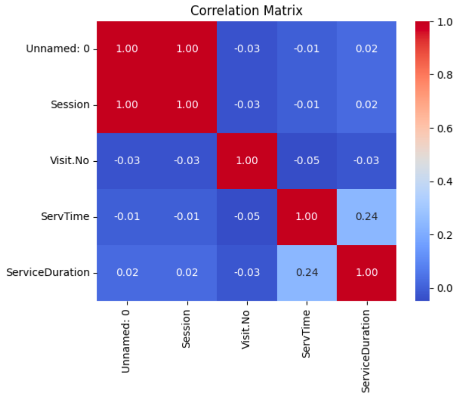
EDA – Service Time (Average)

- Descriptive Statistics

- From the statistical summary, we can see that there are **4,239 records** with an average service duration of approximately **26.59 minutes**. The standard deviation is 34.25 minutes, which indicates a wide-spread in service times. The median value (50%) is 19 minutes, suggesting that half of the services are completed within **19 minutes**.

- The a heatmap showing weak correlation between service time and duration, suggesting other factors influence service duration.





Solution Options

- 1. Process Improvements
    - Review and Optimize Morning Processes:
        - Given the **<u>higher number of outliers</u>** in the morning service times, review the morning procedures. There might be inefficiencies or complex cases that could be better managed or scheduled differently.
    - Standardize Services Where Possible:
        - To reduce the variance in service times, **<u>standardize processes</u>** wherever feasible.
    - Regular Review of Service Time Data:
        - Continuously monitor service times and implement a feedback loop to adjust resources and processes regularly.
    - Employee Training and Incentives:
        - Regular training to enhance the efficiency of staff members, coupled with incentives for improved performance, can motivate staff to manage their time and resources better.

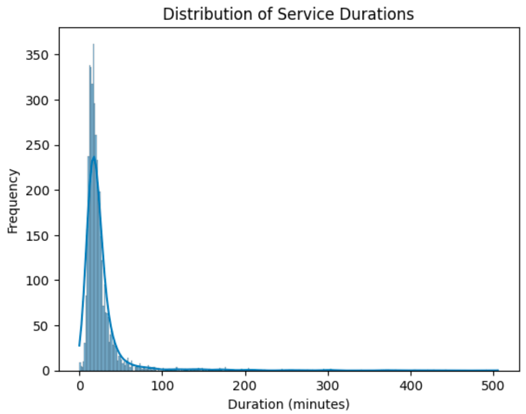
Solution Options

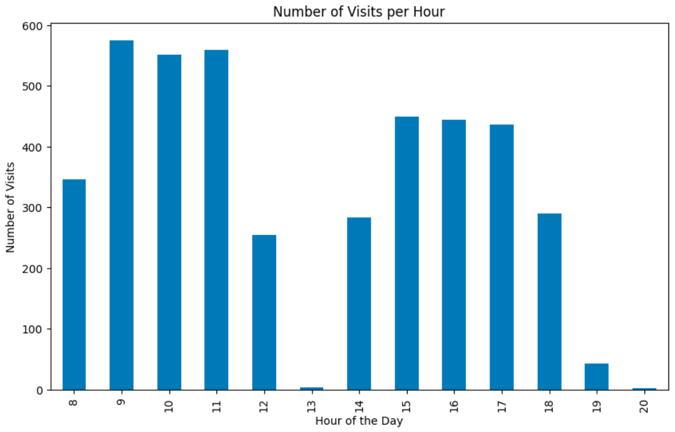
- 2. Data-Driven Decisions
  - Correlation Analysis for Process Improvement:
    - Since the heatmap indicates weak correlation between service time and duration, further investigate **other variables that might influence service duration** (e.g., complexity of services, staff exp., scheduling and waiting times, etc.)
  - Feedback Loops:
    - Implement mechanisms to capture real-time feedback on service duration and customer satisfaction. Use this data to continuously refine processes and training.

EDA – Service Peak (Time)

- The bar chart indicating peak visit times are between 8 AM to 10 AM, with a smaller peak at 2 PM, and a sharp decline after 5 PM.

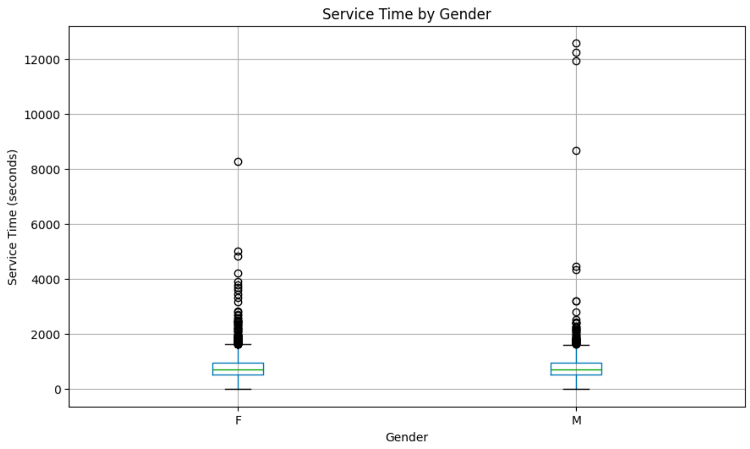
- The boxplot shows a comparison of service durations between **morning and afternoon**. It can be observed that there are more outliers in the morning period compared to the afternoon. This may indicate that certain services in the **morning take significantly longer than average**.





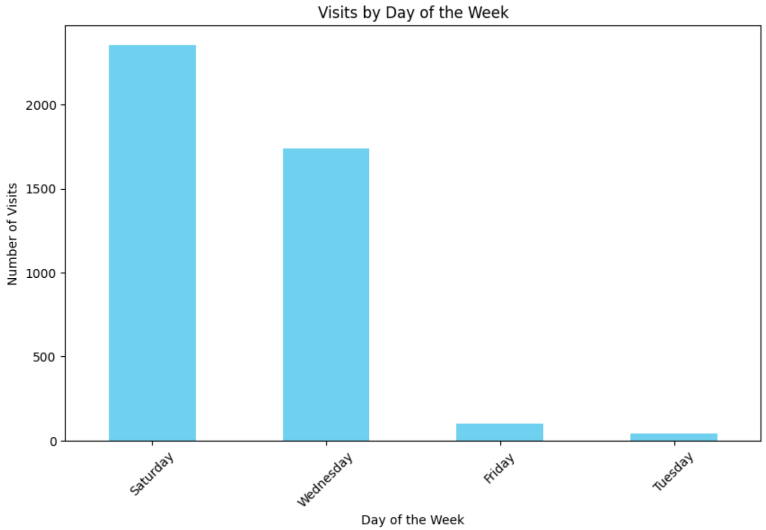
EDA – Service Peak (Gender)

- The boxplot illustrates the distribution of service times separated by gender. The 'F' (Female) and 'M' (Male) categories show a spread of service times, with outliers indicated by points beyond the whiskers. Both distributions have a similar median service time, indicated by the line within the box, but the spread and outliers differ.



EDA – Service Peak (Day)

- The bar chart displays the number of visits for different days of the week. The chart shows that **<u>Sunday and Wednesday</u>** are the busiest days, with Sunday having the most visits. Other days like Friday and Tuesday have significantly fewer visits.



Solution Options

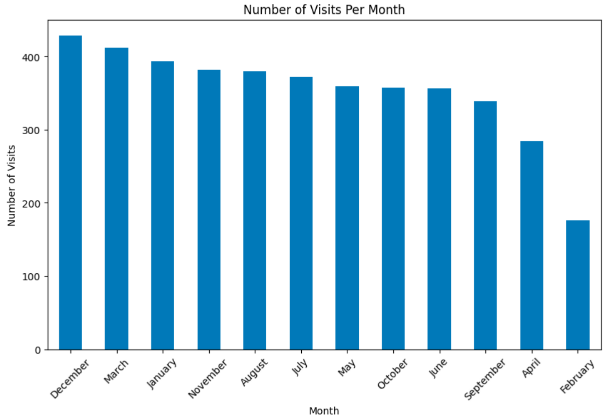
- 3. Staffing Optimization
    - Adjust Staffing Based on Peak Times:
        - Increase staffing during **peak hours (8 AM to 10 AM and at 2 PM)** to handle the high volume of visits efficiently. This should help in reducing wait times and speed up service delivery.
    - Flexible Staffing for Days with Higher Traffic:
        - Allocate more resources on **Sundays and Wednesdays**, which are identified as the busiest days.
        - Consider a slight reduction on days with fewer visits like **Friday and Tuesday**.

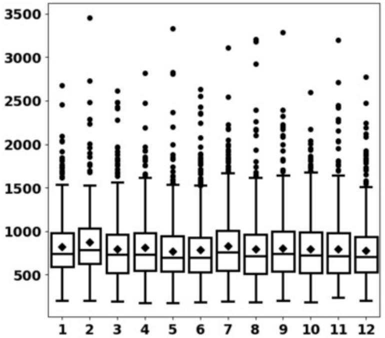
EDA – Service Peak (Month)

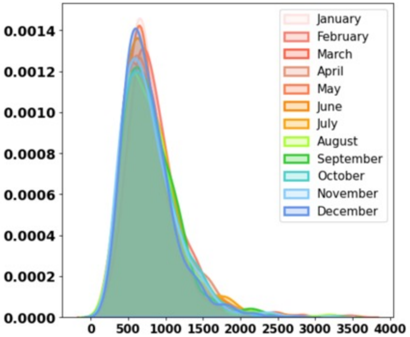
- The bar chart indicates the distribution of visits across months, with **December having the highest visit count** and February the lowest. Seasonal trends, holidays, or weather could explain these variations.

- Boxplot represents some kind of monthly data, possibly service time given the context of the previous image, across 12 categories which are likely to be months (1 to 12). Each boxplot shows the spread of the data with the median, quartiles, and outliers. There is a notable presence of outliers across multiple months.

- Kernel Density Estimate (KDE) plot shows the density of some variable across the same 12 categories. Each month is represented by a different color, and the peaks indicate where data points are concentrated.

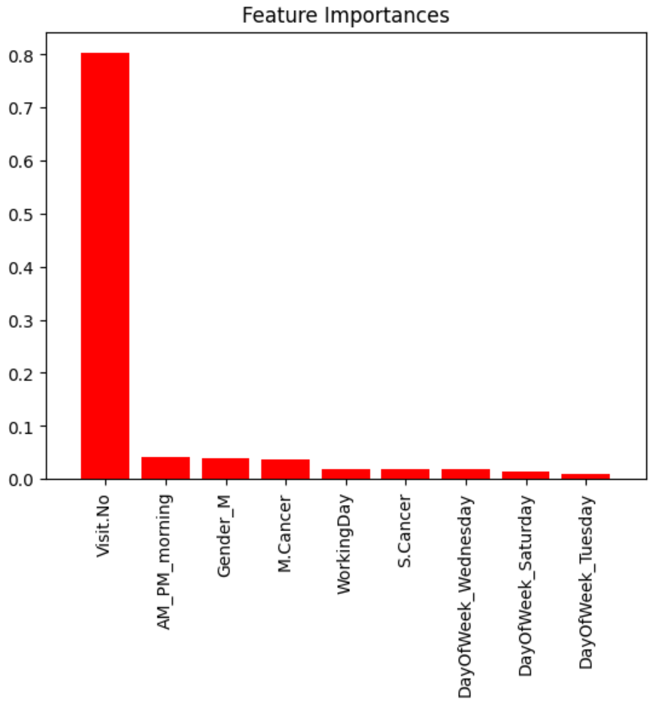






Feature Importance Analysis

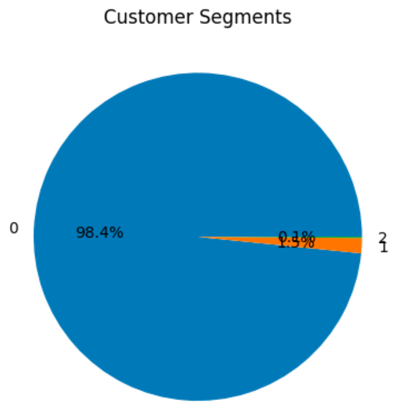
- The bar chat can infer that **<u>"VisitNo" is the most influential feature</u>** for the predictions made by the Random Forest model, and the days of the week seem to have the least influence.
- VisitNo is the most important feature associated with the waiting time issue, which is our objective to improve the operational efficiency.



Model Building

- Feature Engineering
    - A Linear Regression and two ensemble models, Random Forest and Gradient Boosting, are evaluated for **<u>predicting service times</u>**, with Gradient Boosting showing the best performance based on mean squared error.
        - Mean Squared Error: 1134.904231830984
        - Random Forest CV MSE: 1266.6919341222813
        - Gradient Boosting CV MSE: 1212.4904841818957

- Model Enhancement and Model Evaluation with Cross-Validation
    - For customer segmentation, K-Means clustering is applied, resulting in three segments. The pie chart shows that one segment dominates the dataset. Segment means are calculated, indicating **<u>variation in service times across segments</u>**.



Solution Options

- 4. Customer Management
  - Pre-visit Preparations:
    - Encourage or require customers to **fill in necessary forms or perform certain activities** online before their visit. This can significantly reduce processing times per visit.
  - Customer Segmentation and Prioritization:
    - Identify if certain groups require more time and provide appropriate resources or **separate queues** to manage these efficiently.

- 5. Technological Enhancements
  - Implement Advanced Scheduling Tools:
    - Use predictive analytics to forecast busy periods and schedule appointments more effectively. This would help in smoothing spikes in demand and optimizing staff allocation.
  - Queue Management Systems:
    - Deploy digital **queue management systems** to reduce wait times and improve customer flow through the service facility.

03

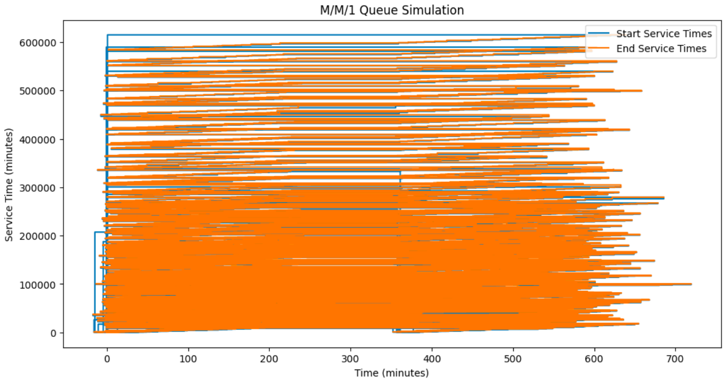
Implementation Plan and Result

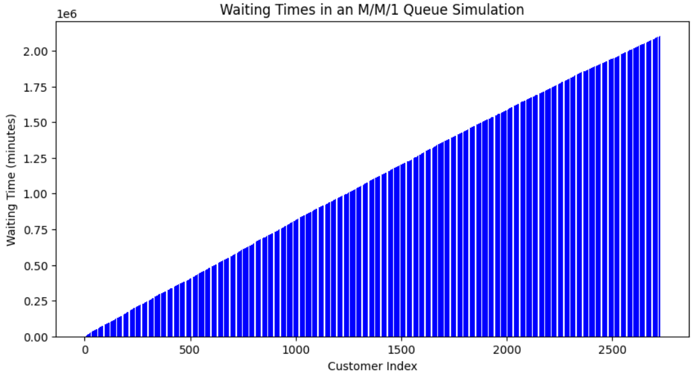
Queueing methods
Sensitivity Analysis

Queue Simulation

Based on previous data analysis, we should focus on **VisitNo** to improve the **ServTime values** and represent the anticipated improvements.

- The visiting clients is a stochastic process. The M/M/1 model in queuing theory is used to analyze systems with exponentially distributed service times. (M: Stands for Memoryless or Markovian/ M: Represents the Poisson arrival process/ 1: Denotes the number of servers in the system) [3, 4]
- As the mean inter-arrival time increases, the average waiting time decreases.
  - When entities arriving less frequently would reduce the congestion and lead to shorter waits.



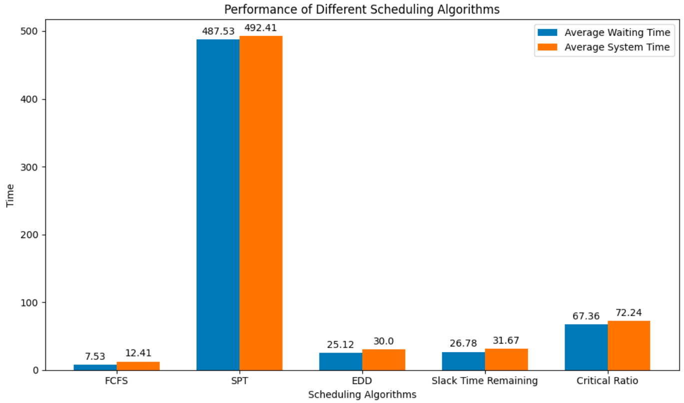


5 Scheduling Methods [5-7]

- 1. First Come, First Served (FCFS)
  - FCFS is typically the simplest and ensures fairness, clients processed in the order of their arrival.
- 2. Shortest Processing Time (SPT)
  - SPT tends to minimize the average waiting time, requiring the least amount of time is selected next.
- 3. Earliest Due Date (EDD)
  - EDD aims at sequenced based on the urgency of their deadlines.
- 4. Slack Time Remaining
  - This rule prioritizes jobs based on the least slack time, which is the time remaining until a job’s due date minus its processing time.
- 5. Critical Ratio
  - The job with the smallest critical ratio (time until due date divided by processing time) is processed first.

Best Method for Scheduling

Scheduling Algorithm	Average Waiting Time	Average System Time
FCFS	7.53	12.41
SPT	487.53	492.41
EDD	25.12	30.00
Slack Time Remaining	26.78	31.67
Critical Ratio	67.36	72.24

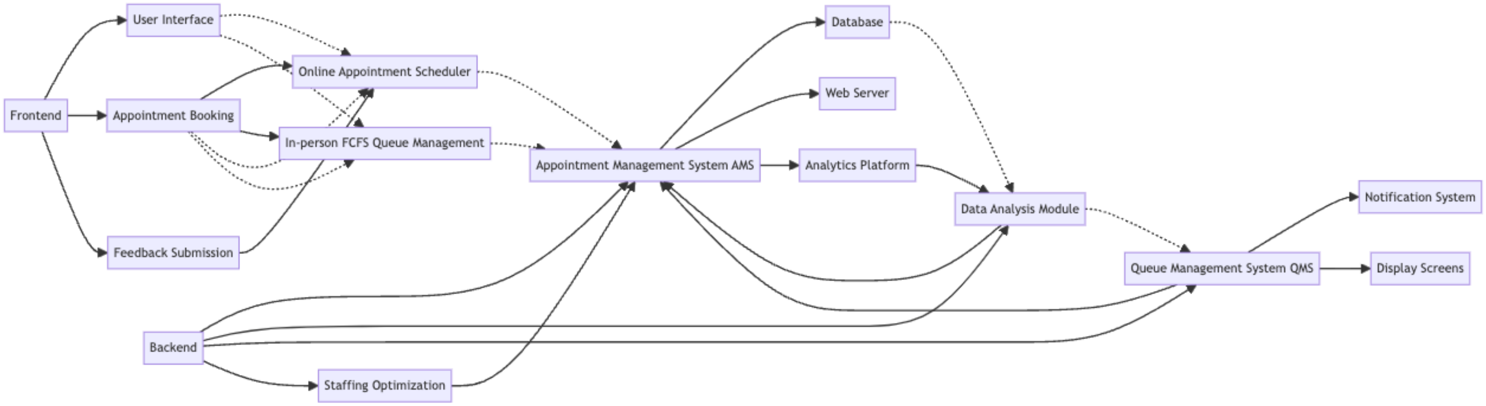


- **In-personal appointment**: FCFS is processing people in the order they arrive.
  - Treatment is uniform and quick.

- **Online appointment**: EDD can effectively manage appointments by prioritizing them based on their scheduled times, ensuring that appointments are handled in a timely manner.
  - Reduction in Waiting Time: By adhering to the scheduled times strictly, the system can minimize delays and reduce overall waiting times for students.

Strategic Plan (combined with solution from data analysis)

- Process improvement: standardization
- Staffing optimization: morning time and Sundays & Wednesdays
- Feedback loop based on data analysis to discover service pattern
- Customer segmentation (online and in person)
- Queue management system



04

Financials, Risks and Contingencies

Operation Plan
Budget Analysis

Management approach

Date/Milestone	Task/Activity	Responsibility
May 2024	Project Kick-off & System Selection	Project Manager, IT Manager
June 2024	System Implementation & Staff Training	IT Team, HR Manager
July 2024	Soft Launch & Marketing Campaign	Operations Manager, Marketing Team
August 2024	Full System Go-Live	Project Team
September 2024	Process Review & Customer Feedback	Operations Manager, Customer Service Team
October 2024	Analysis & Adjustment Phase	All Departments

Financial model

Category	Description	Estimated Cost (USD)	Percentage of Total Budget
Software Development	Custom portal design and development	$100,000	50%
Hardware Acquisition	Servers and infrastructure for hosting the portal	$20,000	10%
Software Licensing	Third-party software for integration (e.g., databases)	$30,000	15%
Testing and Quality Assurance	Ensuring the portal runs smoothly on all devices	$20,000	10%
Training	Training staff to use and manage the portal	$10,000	5%
Marketing and Communication	Promoting the new portal to patients	$10,000	5%
Contingency Fund	Reserved for unexpected expenses	$10,000	5%
Total		**$200,000**	100%

Reference

[1] H. Feng, Y. Jia, S. Zhou, H. Chen, and T. Huang, "A Dataset of Service Time and Related Patient Characteristics from an Outpatient Clinic," Data, vol. 8, no. 3, p. 47, 2023. [Online]. Available: https://doi.org/10.3390/data8030047

[2] "fenghaolin/HanguData," GitHub. [Online]. Available: https://github.com/fenghaolin/HanguData. [Accessed: April 05, 2024].

[3] M. M. Günal and M. Pidd, "Discrete event simulation for performance modelling in health care: a review of the literature," Journal of Simulation, vol. 4, pp. 42-51, 2010.

[4] B. Zeng, A. Turkcan, J. Lin, and M. Lawley, "Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities," Annals of Operations Research, vol. 178, pp. 121-144, Jul. 2010.

[5] A. Kuiper, J. de Mast, and M. Mandjes, "The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice," Omega, vol. 98, p. 102122, 2021.

[6] N. Kortbeek et al., "Designing cyclic appointment schedules for outpatient clinics with scheduled and unscheduled patient arrivals," Performance Evaluation, vol. 80, pp. 5-26, Oct. 2014.

[7] M. R. Hribar et al., "Evaluating and improving an outpatient clinic scheduling template using secondary electronic health record data," in AMIA Annual Symposium Proceedings, vol. 2017, pp. 921, American Medical Informatics Association, 2017.

Thanks!

Any questions?

Email: hin-chi.kwok@berkeley.edu
Website: hinchi-kwok.com