

Project: Temporal Data Analysis

Dataset:

<https://drive.google.com/file/d/15bkGK7PUO1jNlnVxFSC5Pz57rK8kSzmw/view?usp=sharing>

Dataset Description

The dataset contains records of patient visits, including the following fields:

- **ID**: Unique identifier for each patient visit.
- **Session_num**: Numerical identifier indicating the session number of the visit.
- **Month**: Month in which the visit took place.
- **Day**: Day of the week the visit occurred.
- **IsWorkDay**: Boolean indicating if the visit occurred on a working day (**TRUE** or **FALSE**).
- **TimeInDay**: Categorizes the visit time as **morning** or **afternoon**.
- **VisitCount**: The number of times the patient has visited for the same condition.
- **Gender**: Gender of the patient (**F** for female, **M** for male).
- **MalignantCancer**: Boolean indicating if the patient has a history of malignant cancer (**TRUE** or **FALSE**).
- **SkinCancer**: Boolean indicating if the patient has a history of skin cancer (**TRUE** or **FALSE**).
- **TimeStarted**: Timestamp indicating when the visit started (HH:MM:SS).
- **TimePaid**: Timestamp indicating when the payment was completed (HH:MM:SS).
- **Address**: Indicates the patient's address location relative to the city (**In the city**, **Out of city**, **Out of province**, or **NA** for not available).
- **ServiceDurationInSeconds**: Duration of the service provided, in seconds.
-

Task Overview

1. **Data Preparation and Exploration:**
 - Conduct exploratory data analysis to understand the dataset, focusing on temporal patterns related to the day of the week, AM/PM sessions, and service time (ServTime).
 - Visualize the distribution of visit durations (ServTime) across different times of the day and days of the week.
2. **Predictive Modeling:**

- Build a simple regression model to predict the service time (ServTime) based on available features. Consider creating new features that capture temporal aspects (e.g., part of the day, weekday vs. weekend).
 - Discuss your choice of model and features, emphasizing how you incorporated temporal data.
3. **Model Calibration:**
- Implement a basic calibration method (e.g., Platt scaling) on your model's predictions to adjust them to reflect more accurate service times.
 - Evaluate the effectiveness of calibration using a suitable metric (e.g., mean absolute error before and after calibration).

Deliverables

1. **Code Notebook:**
 - A Jupyter notebook containing all codes for EDA, model building, calibration, and evaluation, with comments explaining your steps.
2. **Report:**
 - A brief report within the notebook summarizing:
 - Key insights from your exploratory data analysis.
 - Your approach to building the predictive model, including how temporal data was utilized.
 - The calibration process and its impact on prediction accuracy.

Evaluation Criteria

- **Analytical Skills:** Ability to perform detailed exploratory data analysis and derive meaningful insights from the dataset.
- **Modeling Skills:** Competence in building a predictive model that utilizes temporal data effectively.
- **Calibration Implementation:** Proficiency in applying a calibration method and assessing its impact on improving prediction accuracy.
- **Communication:** Clarity in explaining your analysis, model choice, feature engineering, and findings in both code comments and the summary report.

Instructions for the Candidate

- You are expected to complete this project by Tuesday, 9AM PT
- Document your thought process, any assumptions made, and your reasons for choosing specific models or techniques.
- Ensure your code is well-commented and organized for easy understanding.
- The primary goal is to assess your analytical approach, modeling technique, and ability to work with temporal data, rather than achieving the highest possible accuracy.

This project is designed to gauge your ability to extract insights from temporal data, apply predictive modeling techniques, and understand the importance of model calibration in a real-world context.

Deliverables

1. **Code Notebook:** A Jupyter notebook or similar, containing:
 - EDA with visualizations showing key temporal patterns.
 - Model development process, including feature engineering steps to incorporate temporal information.
 - Implementation of the calibration method and evaluation of its impact.
2. **Brief Report:**
 - A short document or section within the notebook summarizing:
 - Insights from the EDA and how they informed the modeling approach.
 - A description of the model and rationale for the chosen approach to incorporate temporal data.
 - Evaluation of the calibration method's effectiveness and any observed improvements in prediction reliability.