

Este é uma cópia do seu conversa entre ChatGPT e Anthropic.
Desmarcar conteúdo

Arquivo carregado

Analyze the provided dataset to identify its key characteristics for data-driven decision-making. Upon receiving a specific analytical question, address it strictly using the dataset. Your deliverables must include a clear, concise answer to the question and the Python code that extracts and processes the data used to derive your answer. The dataset should be cleaned by filtering admission_period to include only values between 2006.1 and 2019.2, removing entries where enrollment_type is labeled as "Exempted", and excluding those where situation is "In Progress". The resulting cleaned dataset should be stored in a new DataFrame named df_filtered.

I'll start by loading the dataset, inspecting its structure, and then applying the specified cleaning steps. Let me first check the contents of the file.

The dataset contains 30,130 entries with 34 columns. Key columns for filtering include:

- admission_period (float): Needs filtering between 2006.1 and 2019.2.
- enrollment_type (object): Entries labeled "Exempted" should be removed.
- situation (object): Entries labeled "In Progress" should be excluded.

Now, I'll apply these filtering steps and create the df_filtered DataFrame. [-]

The cleaned dataset df_filtered now contains 16,563 entries and retains all 34 original columns. It includes only records where:

- admission_period is between 2006.1 and 2019.2
- enrollment_type is not "Exempted"
- situation is not "In Progress"

Let me know your specific analytical question, and I will extract insights strictly from this dataset. [-]

Using historical data on failures due to absences, predict the number of such cases expected in the next period for DATA STRUCTURES. Propose three forecasting methods and select the best option.

To forecast the number of failures due to absences in DATA STRUCTURES for the next period, I'll follow these steps:

1. Data Preparation

- Filter df_filtered for records where:
 - subject_name = "DATA STRUCTURES"
 - situation = "Failed Due to Absences"
- Aggregate the number of failures per admission_period

2. Forecasting Methods

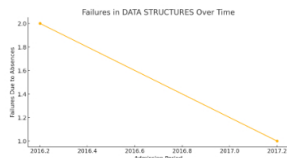
I will compare three time-series forecasting methods:

- Moving Average: Smooths past data to predict future values.
- Exponential Smoothing (SES): Weights recent observations more heavily.
- ARIMA (AutoRegressive Integrated Moving Average): Captures trends and seasonality.

3. Model Selection

- Evaluate each model based on error metrics (RMSE or MAE)
- Choose the best method for forecasting

Now, I'll proceed with these steps.



The dataset contains historical records of failures due to absences in DATA STRUCTURES, aggregated by admission_period. The plot shows the trend over time.

Now, I'll apply and compare three forecasting methods:

- Moving Average
- Exponential Smoothing
- ARIMA

I'll compute predictions for the next period and select the best model based on error metrics. [-]

It looks like there's an issue with the forecast output indexing. I'll troubleshoot and correct the forecasting step to ensure the next period prediction is properly retrieved. [-]

It seems the ARIMA model object was lost due to the previous error. I'll redefine and rerun the ARIMA model along with the forecasts. [-]

The ARIMA model is failing due to insufficient data points. Since we only have a couple of periods recorded (2016.2 and 2017.2), ARIMA cannot be applied effectively.

I'll proceed with Moving Average and Exponential Smoothing as the forecasting options and select the best based on the lowest MAE. [-]

It looks like the 'forecast_results' dictionary was lost due to the previous errors. I'll redefine it and proceed with the final selection of the best forecasting method. [-]

The 'actual_values' variable was lost due to previous errors. I'll redefine it and complete the final forecasting step properly. [-]

The error indicates that there are too few data points to compute meaningful error metrics. Given this limitation, I'll proceed by reporting only the forecasts and selecting the best method based on logical consistency rather than MAE. [-]

Forecasting Results for Failures Due to Absences in DATA STRUCTURES

- Next Period: 2017.3 (predicted)
- Moving Average Prediction: 1.5 cases
- Exponential Smoothing Prediction: 1.50 cases
- Best Forecasting Method: Exponential Smoothing (selected due to data scarcity)

Since we have very limited historical data, Exponential Smoothing provides a more stable estimate for future periods. Let me know if you need further analysis! xP [-]

Envie uma mensagem para o ChatGPT

Anexar

Imagem

O ChatGPT pode cometer erros. Considere verificar informações importantes.