response\_Analysis\_66\_PA - > the document shows the process of achieving the right answer with the usage of prompts and additional context with similar question. The final result has been evaluated as partially answered.

Question: Taking the main variant as basis, what are the main conformance problems?

Initial prompt: (GPT3.5-turbo, temp:0,5, max-length:1200)  
  
SYSTEM: {

*Question ID: 66*

*Question: Taking the main variant as basis, what are the main conformance problems?*

*Prompt: You are a skilled data analyst. Your role is to explain which steps are needed in order to answer the question asked by user. The solution would be written in MS SQL and would query database in order to directly find the answer. Analyze the question and describe in multiply steps how MSSQL query should be built in order to answer the question.*

*The analyst is working on follwing data*

*The event\_log data for the process is stored in the table:BPI\_Challenge\_2019*

*columns:['case\_concept\_name', 'event\_concept\_name', 'event\_time\_timestamp' ]*

*column\_types:['nvarchar', 'nvarchar','datetime2' ].*

*In process mining, a case represents a single instance of a business process.*

*It has a unique ID and attributes like cycle time and costs.*

*Cases are used to analyze and improve processes .*

*ID of a case is stored in case\_concept\_name column.*

*Each case is composed of multiple events.*

*An event name, and identifier, is stored in event\_concept\_name.*

*The time of execution of the event is stored in event\_time\_timestamp.*

*Use minutes in datediff instead of seconds.*

*Use the FOR XML PATH() function to concatenate values instead of STRING\_AGG()*

*If not additionaly specified task duration is a time duration between an event and consecutive event.*

*bottleneck is expectionally long time between two events following each other. For description of bottleneck use event1 -> event2*

*Use 2019-01-17 as today or current date , also use current date for this week, this weekend, today etc.*

*keywords important for answering:*

*['main variant: The main variant represents the most common sequence of activities that occur within a process. \n It can be identified by analyzing the frequency of the different sequences of activities. \n In BPI\_Challenge\_2019, it can be identified by analyzing the frequency of the different sequences of event\_concept\_name within a case (case\_concept\_name column).', '\n', 'conformance problems: Conformance problems occur when there is a mismatch between the actual behavior of a process, as recorded in an event log like BPI\_Challenge\_2019, and the intended or ideal behavior of the process as modeled in a business process model. Conformance checking involves analyzing the event log to identify and diagnose these types of problems.', '\n']*

*QUESTION:*

*Taking the main variant as basis, what are the main conformance problems?*

*Write an answer in following order*

*STEPS:*

*here are the steps to acomplish the task. Do not add additional text or explenation outside of STEPS part. DO not add code outside of CODE part.*

*CODE:*

*```*

*here should be an example MS SQL Query*

*DO NOT USE '--' or any comments in SQL code. Put all comments into description!!*

*```*

*DESCRIPTION:*

*here is a description and reasoning behind the above approach. }*

USER : {

*Taking the main variant as basis, what are the main conformance problems?* }

}

ASSISTANT:  
STEPS:

1. Identify the main variant by finding the most common sequence of events in the cases.

2. Compare the main variant with all other cases to identify any deviations.

3. Calculate the frequency of each deviation and rank them in descending order to find the main conformance problems.

CODE:

```

WITH MainVariant AS (

SELECT case\_concept\_name, STRING\_AGG(event\_concept\_name, '->') WITHIN GROUP (ORDER BY event\_time\_timestamp) AS sequence

FROM BPI\_Challenge\_2019

GROUP BY case\_concept\_name

), VariantFrequency AS (

SELECT sequence, COUNT(\*) AS frequency

FROM MainVariant

GROUP BY sequence

), TopVariant AS (

SELECT TOP 1 sequence

FROM VariantFrequency

ORDER BY frequency DESC

), ConformanceProblems AS (

SELECT mv.case\_concept\_name, mv.sequence, tv.sequence AS main\_variant,

CASE WHEN mv.sequence <> tv.sequence THEN 1 ELSE 0 END AS deviation

FROM MainVariant mv

CROSS JOIN TopVariant tv

), ProblemFrequency AS (

SELECT deviation, COUNT(\*) AS frequency

FROM ConformanceProblems

GROUP BY deviation

)

SELECT deviation, frequency

FROM ProblemFrequency

WHERE deviation = 1

ORDER BY frequency DESC;

```

DESCRIPTION:

The query starts by identifying the main variant by aggregating the event\_concept\_name in the order of event\_time\_timestamp for each case. This is done using the STRING\_AGG() function in the MainVariant CTE. Next, the frequency of each sequence is calculated in the VariantFrequency CTE. The top variant, which is the main variant, is then selected in the TopVariant CTE.

The ConformanceProblems CTE compares the main variant with all other cases to identify deviations. A deviation is marked as 1 if the sequence in a case is different from the main variant, and 0 otherwise. The ProblemFrequency CTE calculates the frequency of each deviation. Finally, the query selects deviations with a value of 1 and orders them by frequency in descending order to identify the main conformance problems.

USER:

Started executing query at Line 1

Msg 9829, Level 16, State 1, Line 1

STRING\_AGG aggregation result exceeded the limit of 8000 bytes. Use LOB types to avoid result truncation.

Total execution time: 00:00:00.991. was a given error. Fix the error.

Do not apologise!!

[as an additional context following code shows the right way to look for main variant: WITH EventSequences AS (

SELECT

case\_concept\_name,

(SELECT event\_concept\_name + '->'

FROM BPI\_Challenge\_2019 b2

WHERE b1.case\_concept\_name = b2.case\_concept\_name

ORDER BY event\_time\_timestamp

FOR XML PATH('')) AS event\_sequence

FROM BPI\_Challenge\_2019 b1

GROUP BY case\_concept\_name

),

SequenceCounts AS (

SELECT event\_sequence, COUNT(\*) AS sequence\_count

FROM EventSequences

GROUP BY event\_sequence

)

SELECT TOP 1 event\_sequence

FROM SequenceCounts

ORDER BY sequence\_count DESC]

ASSISTANT:  
STEPS:

1. Identify the main variant by finding the most common sequence of events in the cases using the FOR XML PATH() function.

2. Compare the main variant with all other cases to identify any deviations.

3. Calculate the frequency of each deviation and rank them in descending order to find the main conformance problems.

CODE:

```

WITH EventSequences AS (

SELECT

case\_concept\_name,

(SELECT event\_concept\_name + '->'

FROM BPI\_Challenge\_2019 b2

WHERE b1.case\_concept\_name = b2.case\_concept\_name

ORDER BY event\_time\_timestamp

FOR XML PATH('')) AS event\_sequence

FROM BPI\_Challenge\_2019 b1

GROUP BY case\_concept\_name

), SequenceCounts AS (

SELECT event\_sequence, COUNT(\*) AS sequence\_count

FROM EventSequences

GROUP BY event\_sequence

), TopVariant AS (

SELECT TOP 1 event\_sequence

FROM SequenceCounts

ORDER BY sequence\_count DESC

), ConformanceProblems AS (

SELECT es.case\_concept\_name, es.event\_sequence, tv.event\_sequence AS main\_variant,

CASE WHEN es.event\_sequence <> tv.event\_sequence THEN 1 ELSE 0 END AS deviation

FROM EventSequences es

CROSS JOIN TopVariant tv

), ProblemFrequency AS (

SELECT deviation, COUNT(\*) AS frequency

FROM ConformanceProblems

GROUP BY deviation

)

SELECT deviation, frequency

FROM ProblemFrequency

WHERE deviation = 1

ORDER BY frequency DESC;

```

DESCRIPTION:

The query starts by identifying the main variant by aggregating the event\_concept\_name in the order of event\_time\_timestamp for each case. This is done using the FOR XML PATH() function in the EventSequences CTE. Next, the frequency of each sequence is calculated in the SequenceCounts CTE. The top variant, which is the main variant, is then selected in the TopVariant CTE.

The ConformanceProblems CTE compares the main variant with all other cases to identify deviations. A deviation is marked as 1 if the sequence in a case is different from the main variant, and 0 otherwise. The ProblemFrequency CTE calculates the frequency of each deviation. Finally, the query selects deviations with a value of 1 and orders them by frequency in descending order to identify the main conformance problems.

SQL SERVER:

A screenshot of a computer

Description automatically generated

The answer has been evaluated as partially answered because only the main deviaition has been returned but not the causes for them.