**Выполнили: Чугин Григорий, Карпов Никита**

**Описание исследования**

Это исследование посвящено сотрудникам, путешествующим с различными целями, поскольку во многих организациях сотрудники путешествуют по работе. Они ездят к клиентам, на конференции или встречи по проектам, которые иногда бывают дорогостоящими. Будучи сотрудником организации, вы не обязаны оплачивать свои собственные командировочные расходы, но компания берет их на себя.

Для поездок внутри страны предварительное разрешение не требуется, т.е. работник может совершить такую поездку и впоследствии потребовать возмещения расходов.

Для международных поездок необходимо разрешение руководителя. Это разрешение можно получить, оформив разрешение на поездку, которое должно быть утверждено до принятия каких-либо мер.

Чтобы получить возмещение расходов на поездку, подается заявление. Это можно сделать сразу после фактической оплаты расходов (например, за перелеты или регистрационные взносы на конференции) или в течение двух месяцев после поездки (например, расходы на гостиницу и питание, которые обычно оплачиваются на месте).

**Описание предоставленных наборов данных**

Набор данных исследования содержит 5 файлов, каждый из которых участвует в исследовании, как показано ниже:

* **Запросы на оплату**(6 886 случаев, 36 796 событий): Этот набор данных содержит все запросы на получение оплаты расходов на поездку.
* **Внутренние декларации** (10 500 случаев, 56 437 событий): Этот набор данных содержит все внутренние поездки, их детали и процедуры.
* Предоплаченные командировочные расходы (2 099 случаев, 18 246 событий):
* **Международные декларации**(6449 случаев, 72151 событие): Этот набор данных содержит все международные поездки, их детали и процедуры.
* **Разрешения на поездки** (7 065 случаев, 86 581 событие): Этот набор данных содержит всю информацию, связанную с разрешениями на поездки (внутренние или международные).

**Задаваемые вопросы**

Следующие вопросы были заданы в задаче, и в ходе следующих шагов мы собираемся ответить на них по мере возможности (поскольку некоторые из них могут не иметь отношения к нашему исследованию, поэтому мы их пропустим):

1. ***Какова пропускная способность туристической декларации от подачи (или закрытия) до оплаты?***
2. Есть ли разница в пропускной способности между национальными и международными поездками?
3. Существуют ли различия между кластерами деклараций, например, между центрами затрат/отделами/проектами и т.д.?
4. Какова пропускная способность на каждом из этапов процесса, т.е. подача заявки, вынесение решения различными ответственными лицами и оплата?
5. ***Где находятся "узкие места" в процессе оформления декларации о поездке?***
6. Где находятся узкие места в процессе получения разрешения на выезд (обратите внимание, что на одно разрешение может приходиться несколько запросов на оплату и деклараций)?
7. ***Сколько туристических деклараций отклоняется на различных этапах обработки и сколько из них никогда не утверждается?***

Затем следуют более подробные вопросы

1. Сколько деклараций о командировках заказывается на проекты?
2. Сколько исправлений было внесено в декларации?
3. Существуют ли двойные платежи?
4. Существуют ли декларации, которым не предшествовало надлежащим образом утвержденное разрешение на поездку? Или даже есть декларации, для которых не существует разрешения?
5. Сколько деклараций о поездке подается самим путешественником и сколько - уполномоченным лицом?
6. Сколько деклараций о поездках сначала отклоняются, потому что они поданы более чем через 2 месяца после окончания поездки, а затем подаются повторно?
7. Отличается ли это между отделами?
8. Сколько деклараций о командировках не утверждается распорядителями бюджета вовремя (7 дней) и затем автоматически перенаправляется руководителям?
9. Наряду с декларациями о поездках существуют также запросы на оплату. Они предназначены для сотрудников, не являющихся сотрудниками TU/e. Есть ли сотрудники TU/e, которые подали запрос на оплату вместо декларации о командировке?

**Шаги к исследованию**

1. **Извлечение журнала:** На этом этапе необработанные журналы входных файлов будут преобразованы в журнал событий путем импорта в pm4py.
2. **Анализ вариантов**: На этом этапе будут распознаны и обсуждены любые различные варианты примера. Таким образом, журнал событий на данном этапе будет преобразован в варианты и их распределения.
3. **Фильтрация**: На этом этапе из набора данных удаляются зашумленные данные и неиспользуемые журналы, чтобы сделать наш анализ более точным. Итак, на входе - журнал событий, а на выходе - отфильтрованный журнал событий.
4. **Обнаружение процессов**: На этом этапе с помощью различных алгоритмов и методов будут продемонстрированы полезные диаграммы (графики). Входом на этом этапе является отфильтрованный журнал событий, а выходом - извлеченная модель.
5. **Проверка соответствия**: На этом этапе оценивается и сравнивается производительность созданных моделей. Входом на этом этапе является модель, созданная на предыдущем этапе, а выходом - важные факторы оценки.
6. **Улучшение процесса**: На этом заключительном этапе обсуждаются возможные улучшения в различных частях процесса. Таким образом, входом здесь является старая модель, а выходом - новая усовершенствованная модель.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Шаги | Вход | Аналитика | Модель | Выход |
| Шаг 1 | Файл | Извлечение | Описательный | Журнал событий |
| Шаг 2 | Журнал событий | Анализ вариантов | Описательный | Вариант |
| Шаг 3 | Рамка данных | Фильтрация | Описательный | Отфильтрованный журнал |
| Шаг 4 | Отфильтрованный журнал | Обнаружение процесса | Предписывающий | Модель |
| Шаг 5 | Модель | Проверка соответствия | Предписывающий | Оценка метрик |

Таблица 1

**Набор данных по международным декларациям**

Первый набор данных, который мы собираемся проанализировать, это международная декларация (имя переменной **idl** в коде), этот набор данных содержит все подробности о международных поездках сотрудников университета TU/e.

Этот журнал имеет больше атрибутов дела, чем его отечественный аналог, но похож на отечественный журнал деклараций, id и **concept:name** идентичны. Кроме того, DeclarationNumber - это id, увеличенный на единицу в 6 189 случаях, в то время как в 260 случаях он UNKNOWN. Кроме того, существуют четыре колонки с идентификаторами, относящимися к соответствующему разрешению, а именно: Номер разрешения на проезд, Номер разрешения на проезд, Идентификатор разрешения и Идентификатор разрешения. Номер разрешения на поездку и номер разрешения на проезд совпадают в 5 970 случаях и различаются в 479 случаях. В отличающихся случаях номер разрешения на проезд Permit 23 раза является UNKNOWN. Идентификатор разрешения и идентификатор разрешения идентичны в 6 001 случае и никогда не являются UNKNOWN. Когда два идентификатора различаются, идентификатор разрешения всегда является разрешением на поездку 423. Кроме того, в 5 970 случаях номер разрешения на поездку представляет собой идентификатор разрешения, увеличенный на единицу, а в 448 случаях - уменьшенный на единицу, и в 31 случае оба номера полностью различаются. Кроме того, существует пять числовых атрибутов дела Amount, RequestedAmount, OriginalAmount, AdjustedAmount и Permit RequestedBudget. Первые три всегда идентичны, тогда как AdjustedAmount отличается только в одном единственном случае, когда сумма 0 корректируется до 100,49. Запрашиваемый бюджет обычно выше, чем сумма. Кроме того, процесс имеет 6 категориальных атрибутов, привязанных к нему, которые закодированы как идентификаторы. Атрибуты включают 6 различных задач, 207 бюджетов, 719 бюджетов разрешений, 825 проектов, 27 организационных единиц и 34 вида деятельности.

**Start activities**

Все события в журнале событий начались со следующих действий, в таблице ниже вы также можете найти их соответствующее распределение:

|  |  |
| --- | --- |
| Declaration SAVED by EMPLOYEE | 8 |
| Declaration SUBMITTED by EMPLOYEE | 407 |
| Permit SUBMITTED by EMPLOYEE | 5294 |
| Start trip | 740 |

Таблица 1

Как видно из таблицы 1, большинство действий начинается с Разрешения, ПОДПИСАННОГО РАБОТНИКОМ, поэтому мы можем рассматривать это действие как хороший вариант для начала потока действий. Однако, мы видим очень низкое количество действий, начатых Декларацией, сохраненной сотрудником, поэтому мы можем рассматривать эту начальную точку как шумные данные и отфильтровать их в следующем разделе.

**Middle activities**

Между начальной и конечной точкой находится 34 вида деятельности, и среди них 15 видов деятельности (как показано в таблице 2) имеют очень низкую частоту, что просто загрязняет набор данных, поэтому на следующих этапах (фильтр) они будут отфильтрованы, чтобы сделать модели ясными.

|  |  |
| --- | --- |
| Declaration REJECTED by PRE\_APPROVER | 84 |
| Declaration SAVED by EMPLOYEE | 75 |
| Declaration REJECTED by MISSING | 103 |
| Permit REJECTED by MISSING | 43 |
| Declaration REJECTED by SUPERVISOR | 126 |
| Declaration APPROVED by SUPERVISOR | 256 |
| Declaration FINAL\_APPROVED by DIRECTOR | 252 |
| Permit REJECTED by PRE\_APPROVER | 25 |
| Permit REJECTED by EMPLOYEE | 231 |
| Declaration REJECTED by DIRECTOR | 4 |
| Permit REJECTED by SUPERVISOR | 92 |
| Permit REJECTED by ADMINISTRATION | 83 |
| Declaration REJECTED by BUDGET OWNER | 40 |
| Permit REJECTED by BUDGET OWNER | 31 |
| Permit REJECTED by DIRECTOR | 1 |

Таблица 2

**End activities**

Все события в журнале событий также завершились следующими действиями, в таблице ниже вы также можете найти их соответствующее распределение:

|  |  |
| --- | --- |
| Declaration SAVED by EMPLOYEE | 54 |
| Declaration REJECTED by EMPLOYEE | 130 |
| Declaration REJECTED by MISSING | 11 |
| End trip | 593 |
| Payment Handled | 5646 |
| Permit REJECTED by MISSING | 8 |

Таблица 2

Некоторые виды деятельности с очень низким числом частот (менее 10) удалены, чтобы сделать таблицу чище. Как вы можете видеть, большинство действий закончилось на Payment Handled, поэтому мы можем рассматривать это действие как хороший вариант в качестве конечной точки. Однако, мы видим, что очень малое количество работ закончилось на Permit REJECTED, поэтому мы можем рассматривать эту конечную точку как шумные данные и отфильтровать их в следующем разделе, так как они не имеют никакого смысла в потоке.

**Анализ вариантов**

Анализ вариантов набора данных показал, что существует 753 различных варианта от начала до конца деятельности, но их распределение можно найти в следующей таблице (для более детального изучения вариантов вы можете найти подробную версию таблицы в файле кода коллаборации).

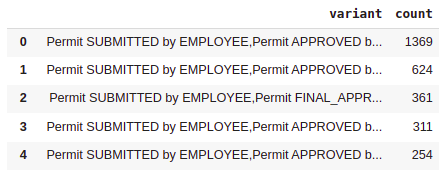


Таблица 3

После построения распределения таблицы 3 на диаграмме с помощью функции логарифма на частотах, мы можем получить следующую диаграмму:

Рис. 1 - Частота вариантов

Как показано на рисунке 1, большинство случаев обрабатывается 5 лучшими вариантами. Таким образом, 5 лучших вариантов могут быть очень полезны в нашем исследовании, и мы будем использовать их в основном на следующих этапах. Однако перед тем, как перейти к следующему шагу, может быть полезно выяснить продолжительность каждого обработанного случая.

## 

Рис. 2 - Длительность рассмотрения дел

На рисунке 2 видно, что максимальная продолжительность рассмотрения дела составляет **742 дня** (это только 1 случай, и он намного больше второго, так что это шум и его можно отфильтровать), а минимальная - **6 дней**, средняя продолжительность рассмотрения дела составляет **86 дней 8 часов**.

**Фильтрация**

Важно, чтобы в наборе данных не было больших шумов и он был пригоден для продолжения работы над следующими разделами и обнаружения процессов. На этом этапе начальные и конечные действия ограничиваются наиболее важными, а действия с очень низкой частотой, которые были распознаны в предыдущих разделах, удаляются из дел, чтобы сделать их как можно более чистыми. В приведенном ниже списке вы можете увидеть отфильтрованные действия для каждого раздела:

* Start Activities: 'Declaration SUBMITTED by EMPLOYEE', 'Permit SUBMITTED by EMPLOYEE', 'Start trip'
* Activities(skipped ones): 'Declaration APPROVED by SUPERVISOR', 'Declaration FINAL\_APPROVED by DIRECTOR', 'Declaration REJECTED by BUDGET OWNER', 'Declaration REJECTED by DIRECTOR', 'Declaration REJECTED by MISSING', 'Declaration REJECTED by PRE\_APPROVER', 'Declaration REJECTED by SUPERVISOR', 'Declaration SAVED by EMPLOYEE', 'Permit REJECTED by ADMINISTRATION', 'Permit REJECTED by BUDGET OWNER', 'Permit REJECTED by DIRECTOR', 'Permit REJECTED by EMPLOYEE', 'Permit REJECTED by MISSING', 'Permit REJECTED by PRE\_APPROVER', 'Permit REJECTED by SUPERVISOR'
* End Activities: 'Payment Handled', 'End trip', 'Declaration REJECTED by EMPLOYEE'

**Process discovery**

Анализ вариантов набора данных показал, что существует 753 различных варианта от начала до конца деятельности, но их распределение можно найти в следующей таблице (для более детального изучения вариантов вы можете найти подробную версию таблицы в файле совместного кода).

Alpha miner

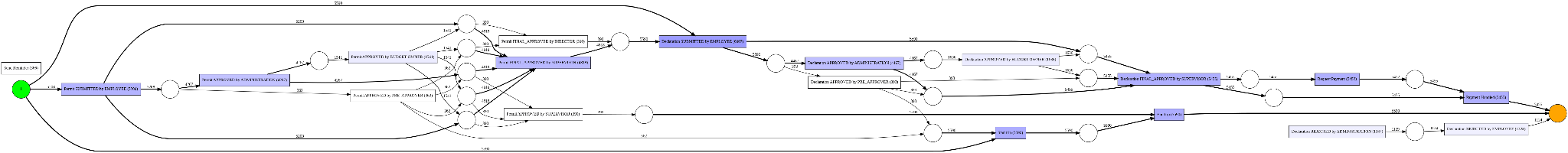


Рис. 3 - Сеть Петри с использованием алгоритма alpha miner

Inductive miner

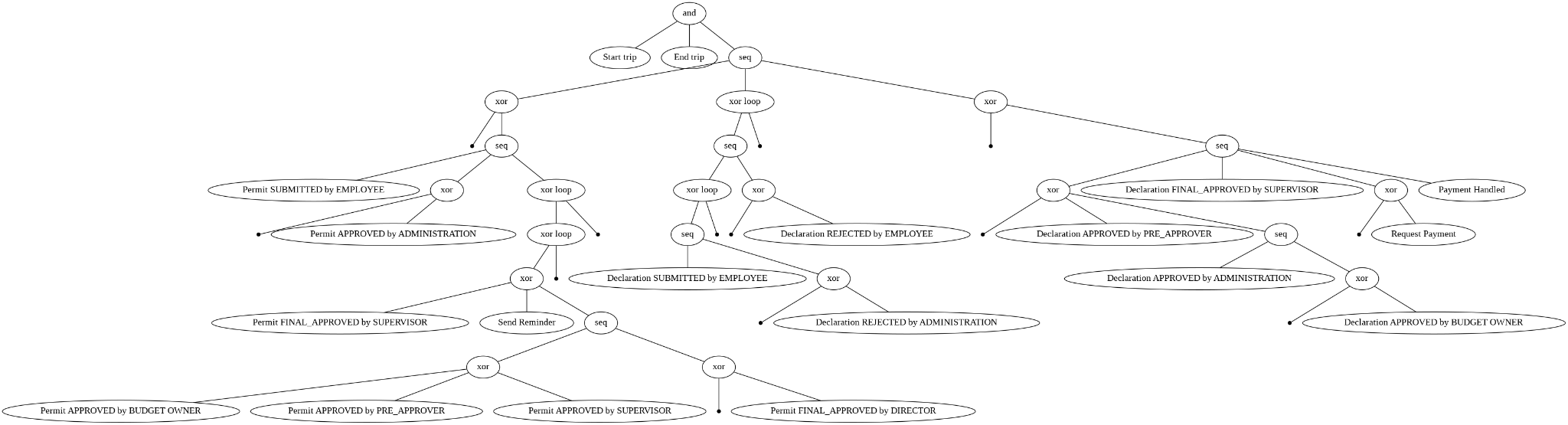


Рис. 4 - дерево с использованием алгоритма inductive miner

Heuristic miner

На следующем графике показан граф с использованием эвристического алгоритма miner.

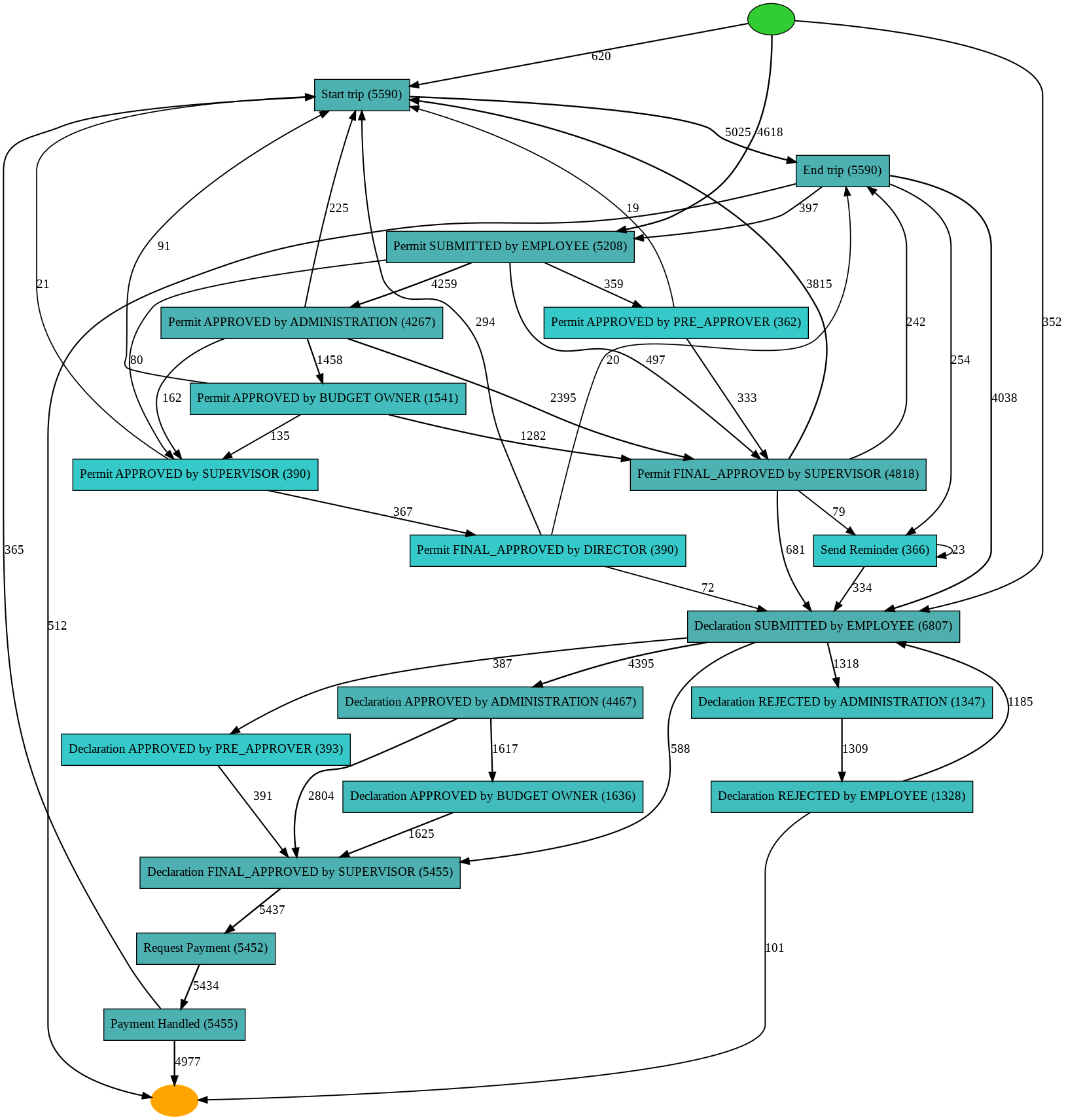


Рис. 6 - График с использованием heuristic miner.

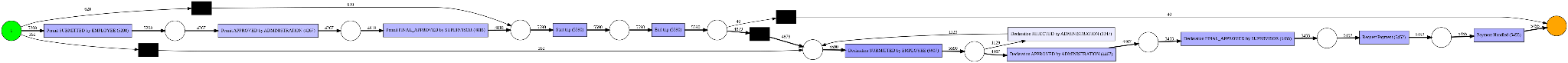


Рис. 7 - Сеть Петри с использованием heuristic miner.

DFG graph

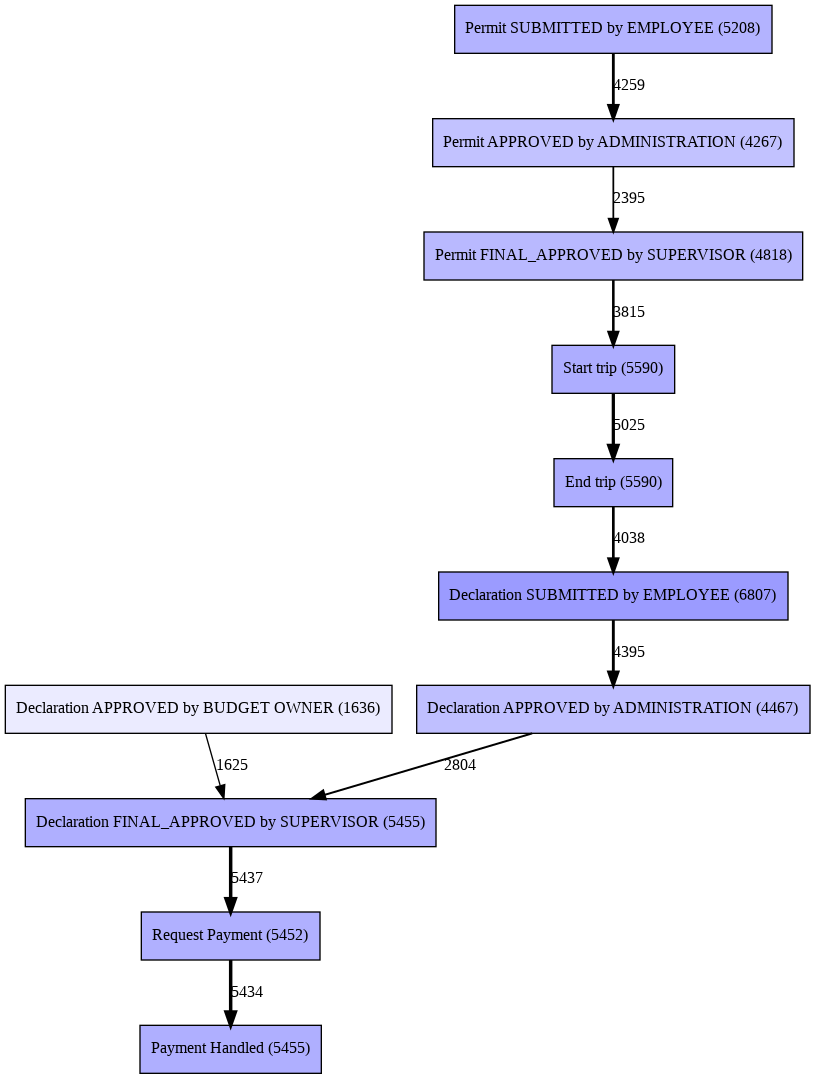


Рис 7 - DFG

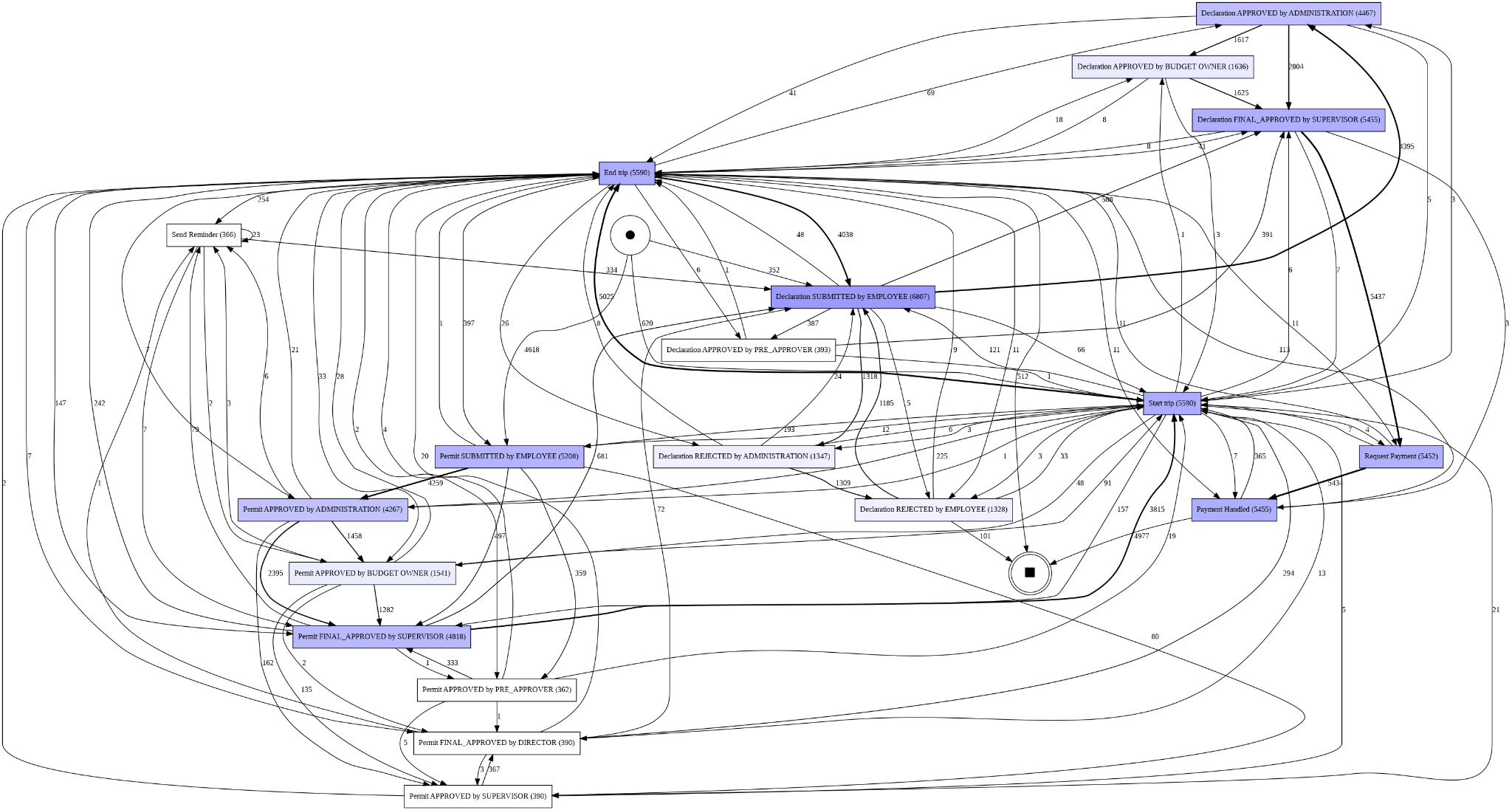


Рис 8 - Полный DFG

### 

**Метрики**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Алгоритм | Precision | Generalization | Simplicity | Fitness |
| Alpha miner | 0.406 | 0.973 | 0.482 | 0.750 |
| Heuristic miner | **0.979** | 0.971 | 0.818 | 0.932 |
| Inductive miner | 0.476 | 0.934 | 0.644 | 1.0 |

Таблица 4

Как показано в таблице 4, существует огромная разница в точности алгоритма heuristic miner, также по сравнению с другими 3 метриками, этот алгоритм имеет лучшую производительность в этом наборе данных.

На рисунке ниже показано воспроизведение токенов эвристической модели:

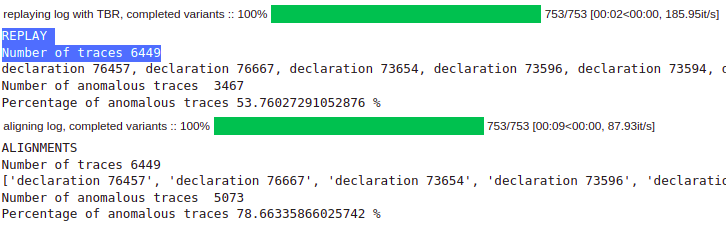


Рис. 9 - проверка соответствия эвристической модели.

**ВОПРОСЫ И ОТВЕТЫ**

**Вопрос 1: Какова пропускная способность декларации о путешествии от подачи (или закрытия) до оплаты?**

Большинство деклараций о поездках были оплачены, и можно рассчитать время прохождения. Что касается международных деклараций, то 95,94 % из 6 449 деклараций были оплачены. Оставшиеся дела либо так и не были утверждены, либо просто сохранены как черновики, которые не были представлены. Чтобы измерить продолжительность рассмотрения одной декларации, рассчитывается среднее значение продолжительности, которое показывает, что запрос от подачи до статуса обработки занимает около **86 дней и 8 часов**.

**Q5: Где находятся "узкие места" в процессе оформления декларации о путешествии?**

Чтобы определить узкие места в процессе оформления декларации о поездке, мы рассчитали продолжительность деятельности по каждому случаю. На основе этого мы рассчитали статистику по всем случаям, включая среднее значение. Мы выбираем виды деятельности с наибольшими узкими местами на основе медианы, а не на основе средней продолжительности, поскольку мы не хотим придавать слишком большое значение выбросам. Кроме того, можно заметить, что действия, которые начинают поездку и заканчивают ее, имеют очень среднюю продолжительность.

**Q7: Сколько туристических деклараций отклоняется на различных этапах обработки и сколько из них никогда не утверждается?**

Чтобы ответить на вопрос 7, мы должны были изучить все действия, которые предполагают отказ или одобрение декларации. Поэтому мы учитывали все действия, включая термины "отклонить" и "одобрить". Для этого мы сначала вычислили матрицу активности случая, в которой подсчитали появление каждого действия для каждого случая. Это позволило нам подсчитать количество случаев отклонения и одобрения. Мы снова проанализировали международные декларации. Для международных деклараций мы рассматривали утверждения и отказы как во время запроса разрешения, так и во время декларирования поездки. Мы снова начали проверять действия, содержащие "REJECT" или "APPROVED", и подсчитали их количество. Затем мы просмотрели все случаи. Из 6 449 случаев, в 5 550 случаях разрешение и декларация были одобрены. В 800 случаях разрешение и декларация не были утверждены. В 279 случаях разрешение было утверждено, но декларация так и не была утверждена, а в 406 случаях декларация была утверждена без утверждения разрешения.

**Выводы**

Наблюдаемый бизнес-процесс был декомпозирован и проанализирован с использованием различных методов поиска процессов. Анализ вариантов и обнаружение процессов позволили нам ответить на вопросы задачи. В международных декларациях было больше разнообразия в деятельности. Множество вариантов было представлено только в одном случае. Этап обнаружения процессов проходил путем создания моделей данных. Анализ соответствия показал лучшие модели для международных деклараций. Лучшей моделью для международных деклараций оказался heuristic miner с **0,93** fitness и **0,97** precision. Хотя точная фильтрация производилась с помощью методов pm4py, оставляя только успешные следы, альфа-майнер показал низкую производительность - **0,75** fitness. alpha miner с уменьшенным количеством следов показал **0,93** fitness. В международных кейсах много различных отказов, поданных несколькими организационными структурами, проектами и другими классами деклараций. Декларации, которые были поданы более чем через 2 месяца после окончания поездки, были отклонены только одним организационным подразделением, что составило **52** процента от общей производительности этого подразделения. Самое важное "узкое место" связано с утверждениями, осуществляемыми различными организационными подразделениями. Утверждение владельцем бюджета занимает дополнительное время для международных поездок, соответственно, около половины и четверти всех утверждений занимают больше времени, чем прямое утверждение администрацией и затем окончательное утверждение руководителем. Вероятно, можно было бы сократить время утверждения на этом этапе, или количество деклараций, отправляемых на утверждение в подразделение-владелец бюджета.