Introduction to Process Mining

```
!pip install -U pm4py
!pip install visualization
```

```
     Downloading PyOpenGL-3.1.0.zip (2.2 MB)
        ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.2/2.2 MB 79.3 MB/s eta 0:00:00
     Preparing metadata (setup.py) ... done
   Collecting xxhash
     Downloading xxhash-3.2.0-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (212 kB)
        ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 212.2/212.2 kB 25.7 MB/s eta 0:00:00
   Collecting rtree
     Downloading Rtree-1.0.1-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.0 MB)
        ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.0/1.0 MB 60.7 MB/s eta 0:00:00
   Requirement already satisfied: shapely in /usr/local/lib/python3.9/dist-packages (from trimesh[easy]->visualization) (2.0.1)
   Requirement already satisfied: lxml in /usr/local/lib/python3.9/dist-packages (from trimesh[easy]->visualization) (4.9.2)
   Collecting pycollada
     Downloading pycollada-0.7.2.tar.gz (107 kB)
        ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 107.6/107.6 kB 9.9 MB/s eta 0:00:00
     Preparing metadata (setup.py) ... done
   Collecting svg.path
     Downloading svg.path-6.2-py2.py3-none-any.whl (40 kB)
        ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 40.9/40.9 kB 4.8 MB/s eta 0:00:00
   Requirement already satisfied: jsonschema in /usr/local/lib/python3.9/dist-packages (from trimesh[easy]->visualization) (4.3.3)
   Requirement already satisfied: requests in /usr/local/lib/python3.9/dist-packages (from trimesh[easy]->visualization) (2.27.1)
   Requirement already satisfied: sympy in /usr/local/lib/python3.9/dist-packages (from trimesh[easy]->visualization) (1.11.1)
   Collecting mapbox-earcut
     Downloading mapbox_earcut-1.0.1-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (105 kB)
        ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 105.2/105.2 kB 13.5 MB/s eta 0:00:00
   Requirement already satisfied: chardet in /usr/local/lib/python3.9/dist-packages (from trimesh[easy]->visualization) (4.0.0)
   Requirement already satisfied: setuptools in /usr/local/lib/python3.9/dist-packages (from trimesh[easy]->visualization) (67.6.1)
   Requirement already satisfied: zipp>=3.1.0 in /usr/local/lib/python3.9/dist-packages (from importlib-resources>=3.2.0->matplotli
   Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.9/dist-packages (from jsonschema->trimesh[easy]->visualiz
   Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in /usr/local/lib/python3.9/dist-packages (from jsc
   Collecting dill>=0.3.6
     Downloading dill-0.3.6-py3-none-any.whl (110 kB)
        ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 110.5/110.5 kB 14.1 MB/s eta 0:00:00
   Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.9/dist-packages (from requests->trimesh[easy]->visua
   Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.9/dist-packages (from requests->trimesh[easy]->vi
   Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.9/dist-packages (from requests->trimesh[easy]->visualizati
   Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.9/dist-packages (from requests->trimesh[easy]
   Collecting ruamel.yaml.clib>=0.2.6
     Downloading ruamel.yaml.clib-0.2.7-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.manylinux_2_24_x86_64.whl (519 kB)
        ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 519.4/519.4 kB 47.3 MB/s eta 0:00:00
   Requirement already satisfied: PyWavelets>=1.1.1 in /usr/local/lib/python3.9/dist-packages (from scikit-image->autolab-core->vis
   Requirement already satisfied: tifffile>=2019.7.26 in /usr/local/lib/python3.9/dist-packages (from scikit-image->autolab-core->v
   Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.9/dist-packages (from scikit-learn->autolab-core->
   Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.9/dist-packages (from sympy->trimesh[easy]->visualization)
   Building wheels for collected packages: PyOpenGL, pycollada
     Building wheel for PyOpenGL (setup.py) ... done
     Created wheel for PyOpenGL: filename=PyOpenGL-3.1.0-py3-none-any.whl size=1745208 sha256=1dd5dba8516efe78b7393dcff2f93737dcff5
     Stored in directory: /root/.cache/pip/wheels/ff/a0/d3/f562b4c97f9b23a5c28dc55a3586a2f5c1d8af766cd8d9bd3a
     Building wheel for pycollada (setup.py) ... done
     Created wheel for pycollada: filename=pycollada-0.7.2-py3-none-any.whl size=127027 sha256=d85ef6637844531c1ddefd8f44dc893617c3
     Stored in directory: /root/.cache/pip/wheels/1c/ab/7d/4974afa70877d57e24a04aa7cddb765c7899793ed5ba515921
   Successfully built PyOpenGL pycollada
   Installing collected packages: svg.path, PyOpenGL, pyglet, xxhash, trimesh, setproctitle, ruamel.yaml.clib, rtree, mapbox-earcut
     Attempting uninstall: PyOpenGL
       Found existing installation: PyOpenGL 3.1.6
       Uninstalling PyOpenGL-3.1.6:
         Successfully uninstalled PyOpenGL-3.1.6
   Successfully installed PyOpenGL-3.1.0 autolab-core-1.1.1 colorlog-6.7.0 dill-0.3.6 freetype-py-2.3.0 mapbox-earcut-1.0.1 multipr
```

Setup

```
!pip install -U pm4py
!pip install visualization
```

```
   Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
   Collecting pm4py
     Downloading pm4py-2.7.2-py3-none-any.whl (1.7 MB)
        ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.7/1.7 MB 60.5 MB/s eta 0:00:00
   Requirement already satisfied: matplotlib in /usr/local/lib/python3.9/dist-packages (from pm4py) (3.7.1)
   Requirement already satisfied: cvxopt in /usr/local/lib/python3.9/dist-packages (from pm4py) (1.3.0)
   Requirement already satisfied: lxml in /usr/local/lib/python3.9/dist-packages (from pm4py) (4.9.2)
   Collecting stringdist
     Downloading StringDist-1.0.9.tar.gz (7.4 kB)
     Preparing metadata (setup.py) ... done
   Requirement already satisfied: graphviz in /usr/local/lib/python3.9/dist-packages (from pm4py) (0.20.1)
   Collecting deprecation
     Downloading deprecation-2.1.0-py2.py3-none-any.whl (11 kB)
```

```
Requirement already satisfied: scipy in /usr/local/lib/python3.9/dist-packages (from pm4py) (1.10.1)
Requirement already satisfied: pydotplus in /usr/local/lib/python3.9/dist-packages (from pm4py) (2.0.2)
Requirement already satisfied: pytz in /usr/local/lib/python3.9/dist-packages (from pm4py) (2022.7.1)
Requirement already satisfied: tqdm in /usr/local/lib/python3.9/dist-packages (from pm4py) (4.65.0)
Requirement already satisfied: networkx in /usr/local/lib/python3.9/dist-packages (from pm4py) (3.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.9/dist-packages (from pm4py) (1.22.4)
Collecting intervaltree
  Downloading intervaltree-3.1.0.tar.gz (32 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: pandas in /usr/local/lib/python3.9/dist-packages (from pm4py) (1.4.4)
Requirement already satisfied: packaging in /usr/local/lib/python3.9/dist-packages (from deprecation->pm4py) (23.0)
Requirement already satisfied: sortedcontainers<3.0,>=2.0 in /usr/local/lib/python3.9/dist-packages (from intervaltree->pm4py) (
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.9/dist-packages (from matplotlib->pm4py) (3.0.9)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.9/dist-packages (from matplotlib->pm4py) (1.4.4)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.9/dist-packages (from matplotlib->pm4py) (8.4.0)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.9/dist-packages (from matplotlib->pm4py) (2.8.2)
Requirement already satisfied: importlib-resources>=3.2.0 in /usr/local/lib/python3.9/dist-packages (from matplotlib->pm4py) (5.
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.9/dist-packages (from matplotlib->pm4py) (4.39.3)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.9/dist-packages (from matplotlib->pm4py) (1.0.7)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.9/dist-packages (from matplotlib->pm4py) (0.11.0)
Requirement already satisfied: zipp>=3.1.0 in /usr/local/lib/python3.9/dist-packages (from importlib-resources>=3.2.0->matplotli
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.9/dist-packages (from python-dateutil>=2.7->matplotlib->pm4py)
Building wheels for collected packages: intervaltree, stringdist
  Building wheel for intervaltree (setup.py) ... done
  Created wheel for intervaltree: filename=intervaltree-3.1.0-py2.py3-none-any.whl size=26114 sha256=afaf0ced199298aca6ed7c4c03f
  Stored in directory: /root/.cache/pip/wheels/ab/fa/1b/75d9a713279796785711bd0bad8334aaace560c0bd28830c8c
  Building wheel for stringdist (setup.py) ... done
  Created wheel for stringdist: filename=StringDist-1.0.9-cp39-cp39-linux_x86_64.whl size=24705 sha256=0221e7123a1a1882a2f23b1e6
  Stored in directory: /root/.cache/pip/wheels/77/24/44/b12b9612fd07e141e6b4f1bb2001c621515bd0cac1014a783a
Successfully built intervaltree stringdist
Installing collected packages: stringdist, intervaltree, deprecation, pm4py
Successfully installed deprecation-2.1.0 intervaltree-3.1.0 pm4py-2.7.2 stringdist-1.0.9
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting visualization
  Downloading visualization-1.0.0-py3-none-any.whl (13 kB)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.9/dist-packages (from visualization) (3.7.1)
Collecting trimesh[easy]
  Downloading trimesh-3.21.4-py3-none-any.whl (680 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 680.8/680.8 KB 37.3 MB/s eta 0:00:00
Collecting pyrender
  Downloading pyrender-0.1.45-py3-none-any.whl (1.2 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.2/1.2 MB 68.2 MB/s eta 0:00:00
```

```
import pandas as pd
from pm4py.objects.conversion.log import converter as log_converter
from pm4py.objects.log.importer.xes import importer as xes_importer
from pm4py.objects.log.util import dataframe_utils

# process mining
from pm4py.algo.discovery.inductive import algorithm as inductive_miner
from pm4py.algo.discovery.alpha import algorithm as alpha_miner
from pm4py.algo.discovery.heuristics import algorithm as heuristics_miner
from pm4py.algo.discovery.dfg import algorithm as dfg_discovery
from pm4py.visualization.dfg import visualizer as dfg_visualization

#sklearn
from sklearn.metrics import pairwise_distances_argmin


# viz

from pm4py.objects.conversion.log import converter as log_converter
from pm4py.algo.discovery.alpha import algorithm as alpha_miner
from pm4py.visualization.petri_net import visualizer as pn_visualizer
from pm4py.visualization.petri_net.util import performance_map
from pm4py.visualization.process_tree import visualizer as pt_visualizer
from pm4py.visualization.heuristics_net import visualizer as hn_visualizer
from pm4py.visualization.process_tree import visualizer as pt_visualizer


# misc
from pm4py.objects.conversion.process_tree import converter as pt_converter

from pm4py.objects.conversion.log import converter as log_converter
from pm4py.objects.log.importer.xes import importer as xes_importer

# process mining
from pm4py.algo.discovery.alpha import algorithm as alpha_miner
from pm4py.algo.discovery.inductive import algorithm as inductive_miner
from pm4py.algo.discovery.heuristics import algorithm as heuristics_miner
from pm4py.algo.discovery.dfg import algorithm as dfg_discovery

# viz
from pm4py.visualization.process_tree import visualizer as pt_visualizer
```

```
from pm4py.visualization.heuristics_net import visualizer as hn_visualizer
from pm4py.visualization.dfg import visualizer as dfg_visualization
```

Loading the data

XES

```
log = xes_importer.apply('running-example.xes')
```

```
/usr/local/lib/python3.9/dist-packages/pm4py/util/dt_parsing/parser.py:76: UserWarning: ISO8601 strings
  warnings.warn(
parsing log, completed traces :: 100%  6/6 [00:00<00:00, 154.22it/s]
```

CSV

```
df = pd.read_csv('running-example.csv')
df = dataframe_utils.convert_timestamp_columns_in_df(df)
df = df.sort_values('time:timestamp')

log = log_converter.apply(df)
```

```
df.sort_values(['case:concept:name', 'time:timestamp']).reset_index(drop=True)
```

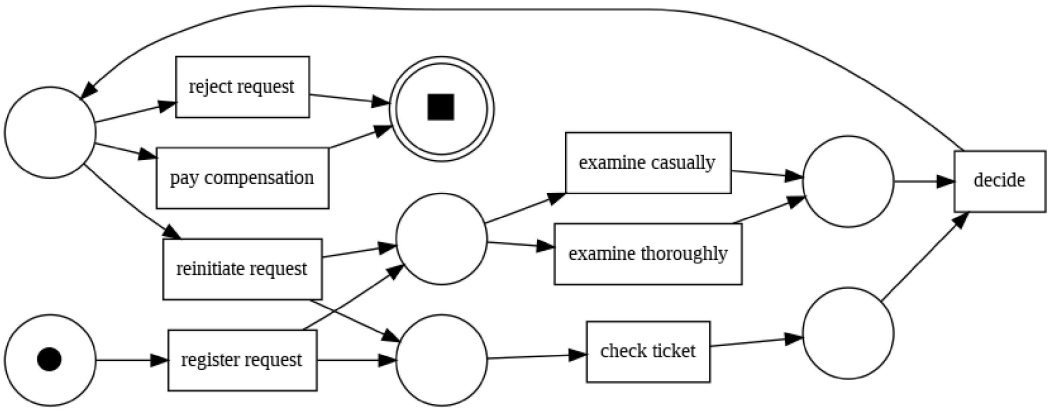
	Activity	Costs	Resource	case:concept:name	case:creator	concept:name	org:resource	time:
0	register request	50	Pete	1	Fluxicon Nitro	register request	Pete	10:0
1	examine thoroughly	400	Sue	1	Fluxicon Nitro	examine thoroughly	Sue	09:0
2	check ticket	100	Mike	1	Fluxicon Nitro	check ticket	Mike	14:1
3	decide	200	Sara	1	Fluxicon Nitro	decide	Sara	10:1
4	reject request	200	Pete	1	Fluxicon Nitro	reject request	Pete	13:2
5	register request	50	Mike	2	Fluxicon Nitro	register request	Mike	10:3
6	check ticket	100	Mike	2	Fluxicon Nitro	check ticket	Mike	11:1
7	examine	400	Sean	2	Fluxicon Nitro	examine	Sean	

Process Mining

Alpha Miner

19	register	50	Pete	4	Fluxicon Nitro	register	Pete	

```
# alpha miner
net, initial_marking, final_marking = alpha_miner.apply(log)

# viz
gviz = pn_visualizer.apply(net, initial_marking, final_marking)
pn_visualizer.view(gviz)
```



20	check ticket	100	Mike	4	Fluxicon Nitro	check ticket	Mike	

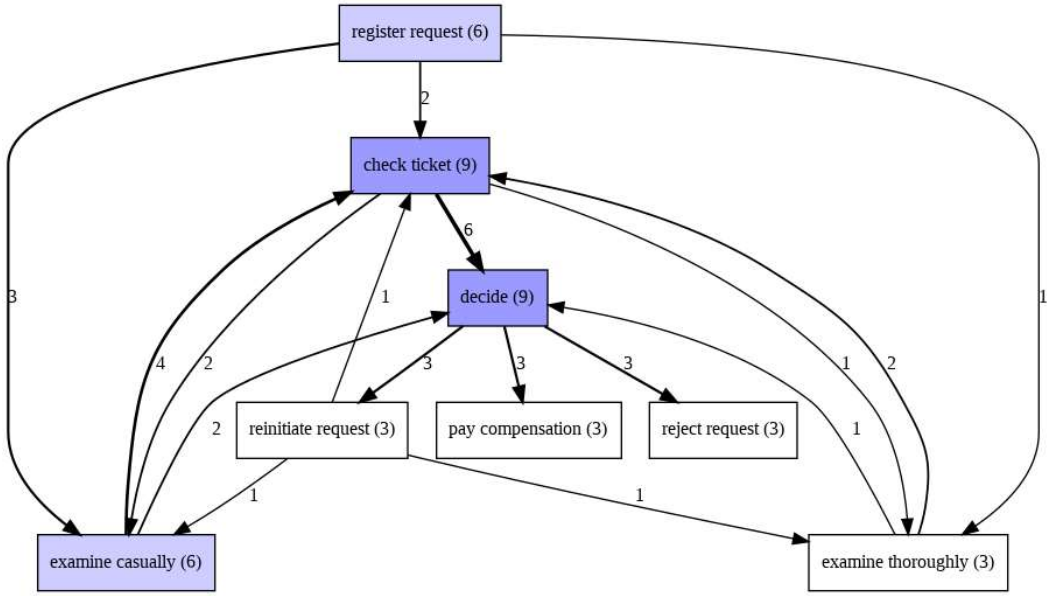
```
# add information about frequency to the viz
parameters = {pn_visualizer.Variants.FREQUENCY.value.Parameters.FORMAT: "png"}
gviz = pn_visualizer.apply(net, initial_marking, final_marking,
                           parameters=parameters,
                           variant=pn_visualizer.Variants.FREQUENCY,
                           log=log)

# save the Petri net
pn_visualizer.save(gviz, "alpha_miner_petri_net.png")
```

replaying log with TBR, completed variants :: 100% 6/6 [00:00<00:00, 137.28it/s]

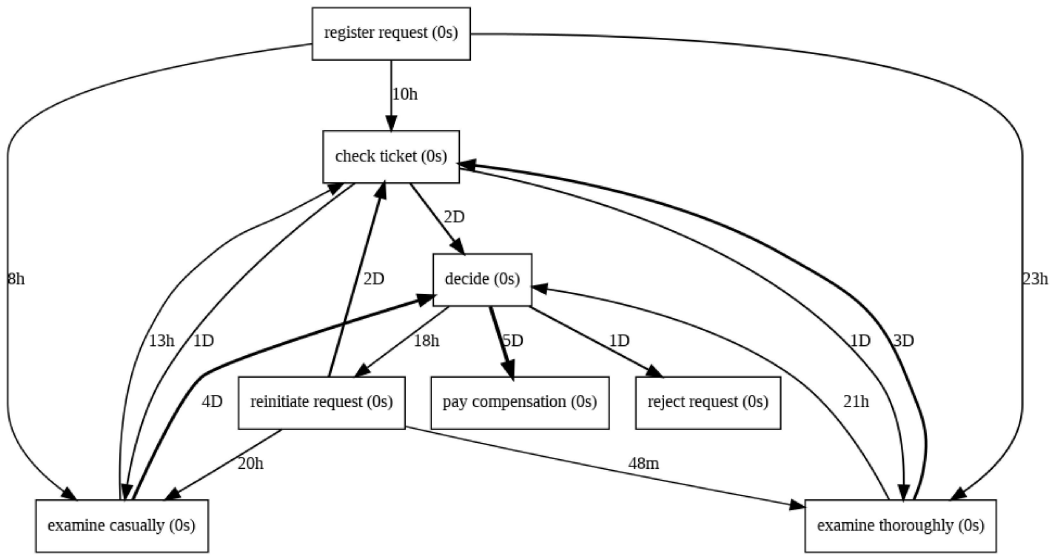
21	casually	400	Mike	5	Fluxicon Nitro	casually	Mike	09:1

Directly-Follows Graph

```
# creatig the graph from log
dfg = dfg_discovery.apply(log)

# viz
gviz = dfg_visualization.apply(dfg, log=log, variant=dfg_visualization.Variants.FREQUENCY)
dfg_visualization.view(gviz)
```

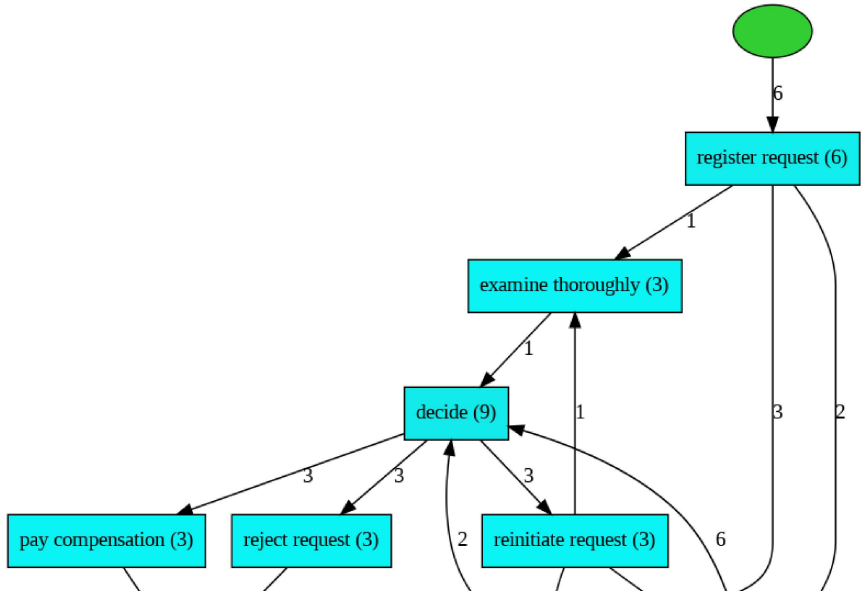



```
# creatig the graph from log
dfg = dfg_discovery.apply(log, variant=dfg_discovery.Variants.PERFORMANCE)

# viz
gviz = dfg_visualization.apply(dfg, log=log, variant=dfg_visualization.Variants.PERFORMANCE)
dfg_visualization.view(gviz)
```

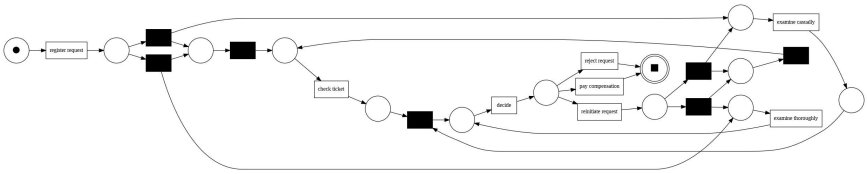


Heuristic Miner

```
# heuristics miner
heu_net = heuristics_miner.apply_heu(log)

# viz
gviz = hn_visualizer.apply(heu_net)
hn_visualizer.view(gviz)
```




```
# heuristics miner
net, im, fm = heuristics_miner.apply(log)

# viz
gviz = pn_visualizer.apply(net, im, fm)
pn_visualizer.view(gviz)
```

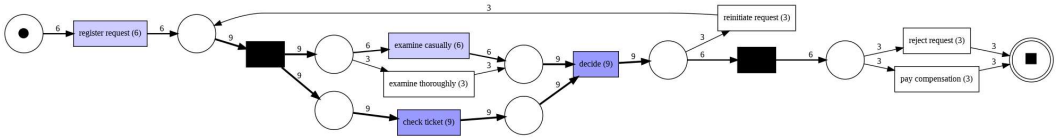


Inductive Miner

```
# create the process tree
tree = inductive_miner.apply(log)

# viz
gviz = pt_visualizer.apply(tree)
pt_visualizer.view(gviz)
```


```
# convert the process tree to a petri net
net, initial_marking, final_marking = pt_converter.apply(tree)

# alternatively, use the inductive_miner to create a petri net from scratch
# net, initial_marking, final_marking = inductive_miner.apply(log)

# viz
parameters = {pn_visualizer.Variants.FREQUENCY.value.Parameters.FORMAT: "png"}
gviz = pn_visualizer.apply(net, initial_marking, final_marking,
                           parameters=parameters,
                           variant=pn_visualizer.Variants.FREQUENCY,
                           log=log)
pn_visualizer.view(gviz)
```

replaying log with TBR, completed variants :: 100% 6/6 [00:00<00:00, 176.02it/s]



Colab paid products - Cancel contracts here

✓ 0s completed at 10:03 PM