```
!pip uninstall CuPy
```

```
WARNING: Skipping CuPy as it is not installed.
```

Чтение данных

```
import pandas as pd
import pm4py

file_path = '/content/final_dataset.csv'
pd_df = pd.read_csv(file_path)

df = pm4py.format_dataframe(
    pd_df,
    case_id='case concept:name',
    activity_key='event concept:name',
    timestamp_key='event time:timestamp',
    timest_format='%d-%m-%Y %H:%M:%S.%f'
)

df.head()
```

```
{"type":"dataframe","variable_name":"df"}
```

Кластера

```
from sklearn.preprocessing import StandardScaler
event_log = pm4py.convert_to_event_log(df)

df['duration'] = (df['event time:timestamp'] - df.groupby('case
concept:name')['event
time:timestamp'].shift()).fillna(pd.Timedelta(seconds=0))
df['duration'] = df['duration'].dt.total_seconds()

features = df[['duration']]

scaler = StandardScaler()
data_scaled = scaler.fit_transform(features)

kmeans = KMeans(n_clusters=3, random_state=0)
clusters = kmeans.fit_predict(data_scaled)

df['cluster'] = clusters
df.head()
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
```

```
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(
Exception ignored on calling ctypes callback function: <function
ThreadpoolController._find_libraries_with_dl_iterate_phdr.<locals>.mat
ch_library_callback at 0x7d41071f24d0>
Traceback (most recent call last):
  File "/usr/local/lib/python3.10/dist-packages/threadpoolctl.py",
line 1005, in match_library_callback
    self._make_controller_from_path(filepath)
  File "/usr/local/lib/python3.10/dist-packages/threadpoolctl.py",
line 1175, in _make_controller_from_path
    lib_controller = controller_class(
  File "/usr/local/lib/python3.10/dist-packages/threadpoolctl.py",
line 114, in __init__
    self.dynlib = ctypes.CDLL(filepath, mode=_RTLD_NOLOAD)
  File "/usr/lib/python3.10/ctypes/__init__.py", line 374, in __init__
    self._handle = _dlopen(self._name, mode)
OSError:
/usr/local/lib/python3.10/dist-packages/cvxopt.libs/libopenblasp-r0-
5c2b7639.3.23.so: cannot open shared object file: No such file or
directory
```

{"type":"dataframe","variable_name":"df"}

Оценка качества

```
from sklearn.metrics import silhouette_score
import pm4py
from pm4py.algo.evaluation.replay_fitness import algorithm as
replay_fitness

# Silhouette Score
silhouette_avg = silhouette_score(data_scaled, clusters)
```

```
---------------------------------------------------------------------
-----
TypeError                                 Traceback (most recent call
last)
<ipython-input-24-e35905a18c38> in <cell line: 9>()
      7
      8 # Fitness
----> 9 fitness_result = replay_fitness.evaluate(event_log)
     10
     11 # Вывод результатов

/usr/local/lib/python3.10/dist-packages/pm4py/algo/evaluation/replay_f
itness/algorithm.py in evaluate(results, parameters, variant)
```

```
    117         Fitness evaluation
    118     """
--> 119     return exec_utils.get_variant(variant).evaluate(results,
parameters=parameters)

/usr/local/lib/python3.10/dist-packages/pm4py/algo/evaluation/replay_f
itness/variants/token_replay.py in evaluate(aligned_traces,
parameters)
     57         parameters = {}
     58     no_traces = len(aligned_traces)
---> 59     fit_traces = len([x for x in aligned_traces if
x["trace_is_fit"]])
     60     sum_of_fitness = sum([x["trace_fitness"] for x in
aligned_traces])
     61     perc_fit_traces = 0.0

/usr/local/lib/python3.10/dist-packages/pm4py/algo/evaluation/replay_f
itness/variants/token_replay.py in <listcomp>(.0)
     57         parameters = {}
     58     no_traces = len(aligned_traces)
---> 59     fit_traces = len([x for x in aligned_traces if
x["trace_is_fit"]])
     60     sum_of_fitness = sum([x["trace_fitness"] for x in
aligned_traces])
     61     perc_fit_traces = 0.0

/usr/local/lib/python3.10/dist-packages/pm4py/objects/log/obj.py in
__getitem__(self, key)
    235
    236     def __getitem__(self, key):
--> 237         return self._list[key]
    238
    239     def __iter__(self):

TypeError: list indices must be integers or slices, not str
```

```
silhouette_avg
```

```
0.8292852292776111
```

```
from pm4py.algo.conformance.tokenreplay import algorithm as
token_replay
from pm4py.algo.evaluation.replay_fitness import algorithm as
replay_fitness

net, initial_marking, final_marking =
pm4py.discover_petri_net_alpha(event_log)

aligned_traces = token_replay.apply(event_log, net, initial_marking,
final_marking)
```

```
# Оценка метрики Fitness
fitness_result = replay_fitness.evaluate(aligned_traces)
```

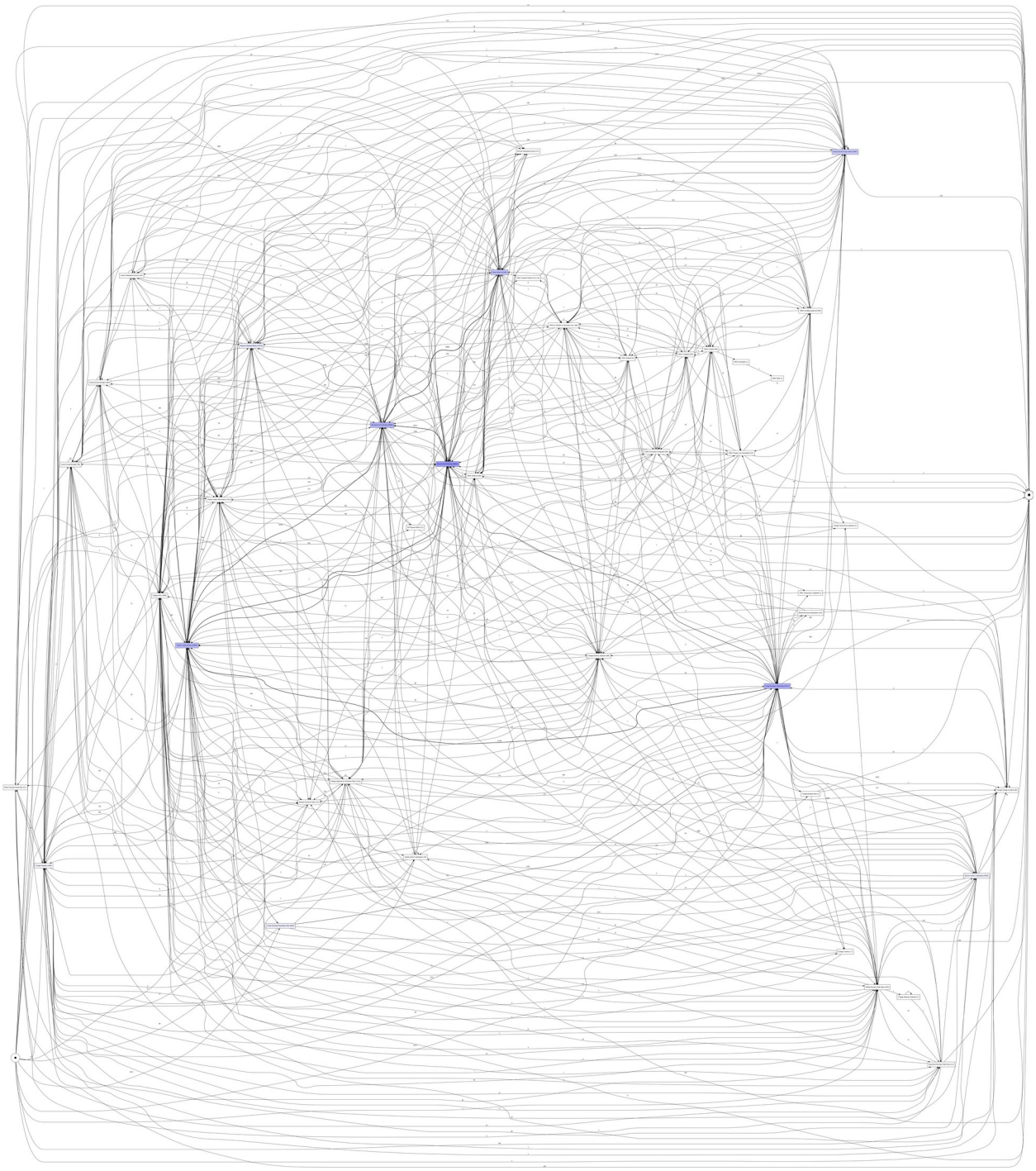
{"model_id":"ea05a044daa44d2dba623803c69f34cd","version_major":2,"version_minor":0}

```
fitness_result
```

```
{'perc_fit_traces': 0.0,
 'average_trace_fitness': 0.32459031290543233,
 'log_fitness': 0.28022625410597896,
 'percentage_of_fitting_traces': 0.0}
```

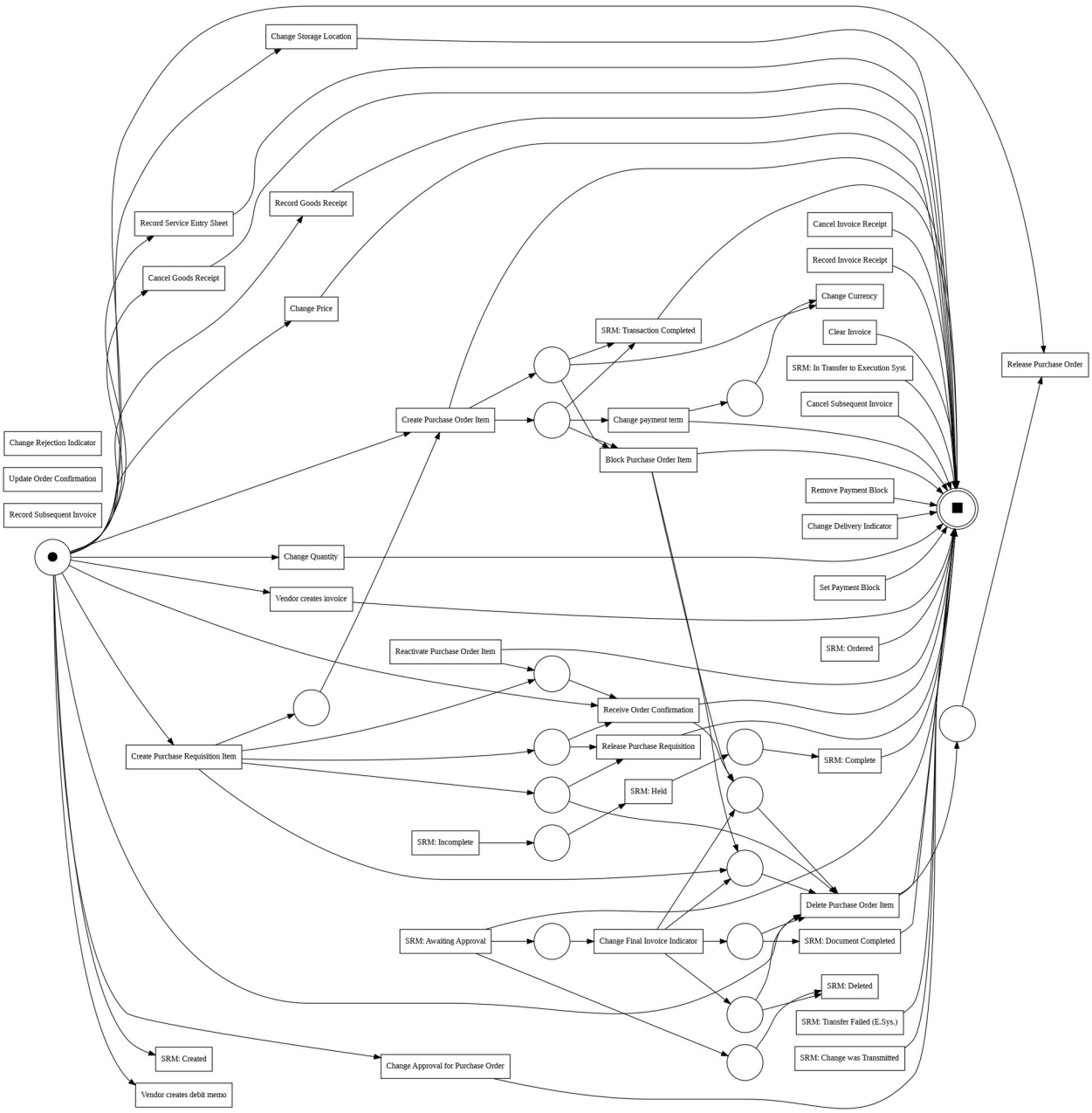
DFG визуализация

```
from pm4py.discovery import discover_dfg

dfg = discover_dfg(event_log)

pm4py.view_dfg(dfg[0], dfg[1], dfg[2])
```



Сеть Петри

```
net, im, fm = pm4py.discover_petri_net_alpha(event_log)

pm4py.view_petri_net(net, im, fm)
```



Среднее время работы

```
df['duration'] = (df['event time:timestamp'] - df.groupby('case
concept:name')['event
time:timestamp'].shift()).fillna(pd.Timedelta(seconds=0))
df['duration'] = df['duration'].dt.total_seconds()

features = df[['duration']]

scaler = StandardScaler()
```

```
data_scaled = scaler.fit_transform(features)

kmeans = KMeans(n_clusters=3, random_state=0)
clusters = kmeans.fit_predict(data_scaled)

df['cluster'] = clusters

avg_time_per_cluster = df.groupby('cluster')['duration'].mean()
avg_time_per_cluster
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/
_kmeans.py:870: FutureWarning: The default value of `n_init` will
change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  warnings.warn(

cluster
0    4.371160e+05
1    3.594901e+08
2    6.022823e+06
Name: duration, dtype: float64
```