

# Cas d'étude process mining

*Étude des logs de rentes et des discrimination dans le processus associé*

---

<b>Introduction</b>	<b>1</b>
<b>Analyses</b>	<b>1</b>
Analyse globale des fichiers	1
Analyse des discrimination	13
Statistiques générales	14
Sexisme	15
Âgisme	19
Discrimination par langue parlée	23
Discrimination par status marital	25
Discrimination par citoyenneté	27
Discrimination par "protected"	29
Discrimination par années d'études	31
Récapitulatif	36
<b>Conclusion</b>	<b>36</b>

---

---

## Introduction

Nous nous intéressons dans ce sujet à l'étude des events logs renting. L'ensemble de nos travaux sont consultables sur [GitHub](#).

Le rendu comprend également [analysis.ipynb](#), le notebook de développement sur lequel nous avons fait nos analyses générales des logs. Un autre notebook nommé [discrimination\\_analysis.ipynb](#) permet de faire l'analyse de la discrimination par divers facteurs.

Vous trouverez également dans ce rendu un classeur consultable [ici](#) où se trouve un résumé des statistiques avec des graphiques pour illustrer nos résultats.

Les fichiers que nous étudions sont renting\_log\_XXX.xes ils sont disponibles dans le dépôt [GitHub](#) dans le dossier data.

## Analyses

### Analyse globale des fichiers

Statistiques générales

Nombre de cas / traces : 105555

Nombre de variantes : 610

Nombre moyen d'activités par cas : 0.0019800104211074795

Nombre minimal d'activités par cas : 1

Nombre maximal d'activités par cas : 21

Durée minimale : 79856.0

Durée maximale : 5279770.0

Durée moyenne : 704362.7504

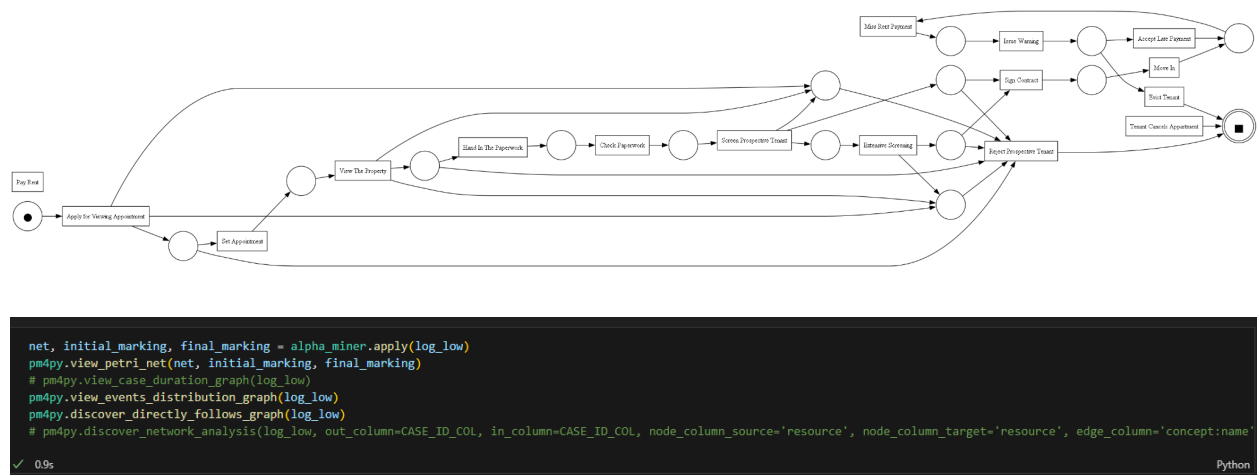


Figure 1 et 2 : PN découvert avec l'algorithme alpha miner et le code associé

Afin de nous figurer le processus nous avons appliqué l'algorithme alpha miner qui nous a retourné la figure ci-dessus.

Ensuite nous pouvons nous figurer les noms des events et des concepts dans l'event log avec ce code:

```

log_low.columns

```

```

Index(['activity', 'resource', 'time', 'concept:name', 'time:timestamp',
      '@@index', 'case:concept:name', 'case:german speaking',
      'case:yearsOfEducation', 'case:age', 'case:gender', 'case:citizen',
      'case:protected', 'case:married', 'case:@@case_index', 'case:case'],
      dtype='object')

```

Figure 3 : Liste des event et concepts présent dans le log low

Et nous pouvons aussi obtenir la répartition des événements dans un log avec le code vus dans la figure 2.

```
(({'Accept Late Payment', 'Miss Rent Payment'): 37,
 ('Accept Late Payment', 'Pay Rent'): 894,
 ('Apply for Viewing Appointment', 'Reject Prospective Tenant'): 643,
 ('Apply for Viewing Appointment', 'Set Appointment'): 9357,
 ('Check Paperwork', 'Screen Prospective Tenant'): 8049,
 ('Extensive Screening', 'Reject Prospective Tenant'): 1592,
 ('Extensive Screening', 'Sign Contract'): 1093,
 ('Hand In The Paperwork', 'Check Paperwork'): 8049,
 ('Issue Warning', 'Accept Late Payment'): 931,
 ('Issue Warning', 'Evict Tenant'): 137,
 ('Miss Rent Payment', 'Issue Warning'): 1068,
 ('Move In', 'Miss Rent Payment'): 165,
 ('Move In', 'Pay Rent'): 3245,
 ('Pay Rent', 'Miss Rent Payment'): 866,
 ('Pay Rent', 'Pay Rent'): 16868,
 ('Pay Rent', 'Tenant Cancels Appartment'): 3273,
 ('Screen Prospective Tenant', 'Extensive Screening'): 2685,
 ('Screen Prospective Tenant', 'Reject Prospective Tenant'): 3047,
 ('Screen Prospective Tenant', 'Sign Contract'): 2317,
 ('Set Appointment', 'View The Property'): 9357,
 ('Sign Contract', 'Move In'): 3410,
 ('View The Property', 'Hand In The Paperwork'): 8049,
 ('View The Property', 'Reject Prospective Tenant'): 1308},
 {'Apply for Viewing Appointment': 10000},
 {'Reject Prospective Tenant': 6590,
 'Tenant Cancels Appartment': 3273,
 'Evict Tenant': 137})
```

Figure 4 : Liste des events et nombre d'occurence dans le log renting low

En outre nous pouvons dessiner le graph de fréquence des events

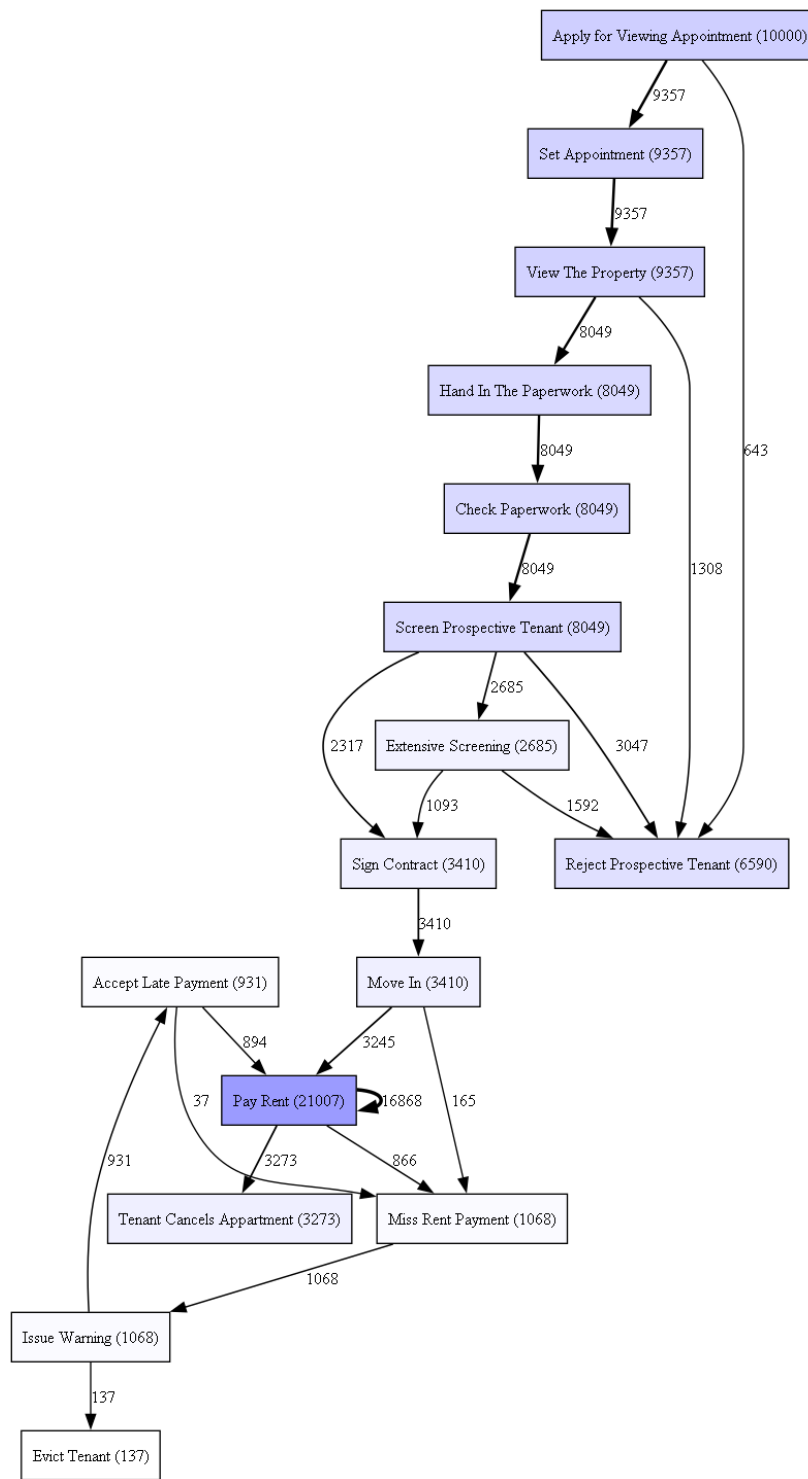
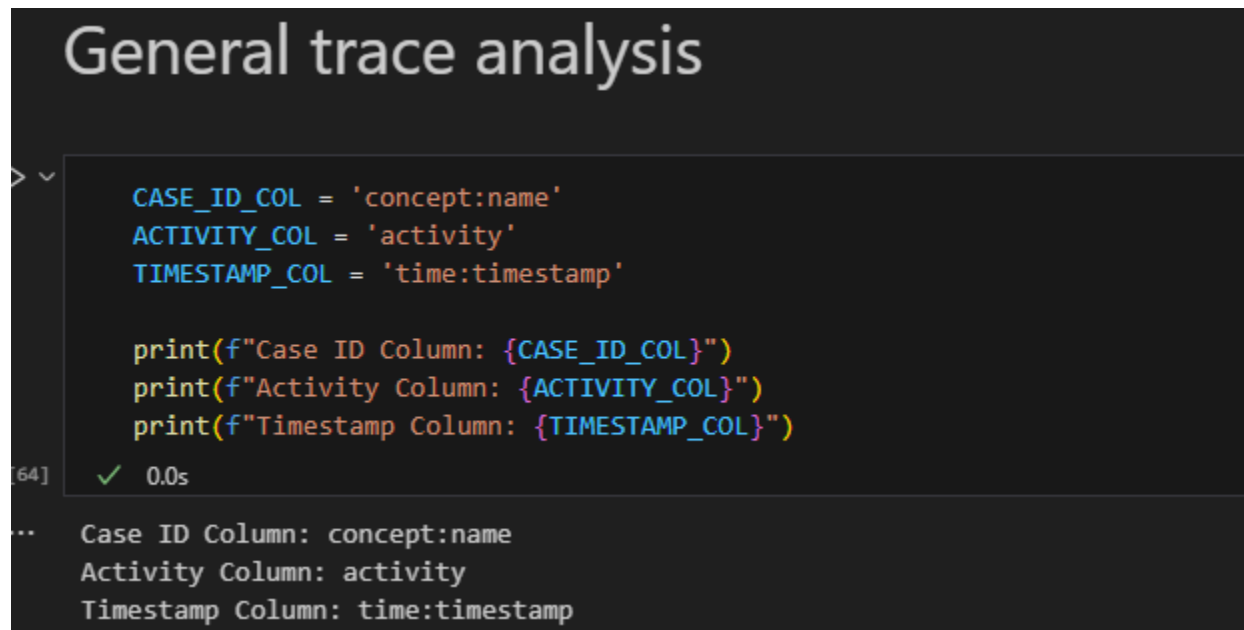


Figure 5 : Liste des events et nombre d'occurrence dans le log renting low

Lors d'une analyse de log il est important de connaître le nom des différents éléments comme le case id, la colonne d'activité et de timestamp. Nous les trouvons dans notre cas avec le code suivant.



```
General trace analysis

> ✓ 0.0s

CASE_ID_COL = 'concept:name'
ACTIVITY_COL = 'activity'
TIMESTAMP_COL = 'time:timestamp'

print(f"Case ID Column: {CASE_ID_COL}")
print(f"Activity Column: {ACTIVITY_COL}")
print(f"Timestamp Column: {TIMESTAMP_COL}")

... Case ID Column: concept:name
Activity Column: activity
Timestamp Column: time:timestamp
```

Figure 6 : Les colones

Nous pouvons également trouver des statistiques tel que le nombre de variant, d'événement avec ce fragment de code.

Statistics for the `low` dataset:

```
num_cases = len(log_low.groupby(CASE_ID_COL))
num_events = len(log_low)

variants = pm4py.get_variants(log_low)
num_variants = len(variants)

print(f"Number of cases: {num_cases}")
print(f"Number of events: {num_events}")
print(f"Number of variants: {num_variants}")
```

✓ 0.0s

```
Number of cases: 16
Number of events: 96440
Number of variants: 508
```

Figure 7 : Statistiques générales

D'autres statistiques intéressantes sont les event de début et de fin de traces que nous pouvons trouver grâce à ce fragment de code. Ou nous y constatons par log le nom de l'event et le nombre de fois qu'il apparaît en début / fin de trace. Nous voyons par exemple dans les exemples ci-contre que toutes les traces commencent par "Apply for Viewing Appointment". Mais les event de fin sont plus variés.

```

#Getting first and last event
first_event = pm4py.get_start_activities(log_low)
last_event = pm4py.get_end_activities(log_low)

print(f"First event: {first_event}")
print(f"Last event: {last_event}")

#Getting first and last event
first_event = pm4py.get_start_activities(log_medium)
last_event = pm4py.get_end_activities(log_medium)

print(f"First event: {first_event}")
print(f"Last event: {last_event}")

#Getting first and last event
first_event = pm4py.get_start_activities(log_high)
last_event = pm4py.get_end_activities(log_high)

print(f"First event: {first_event}")
print(f"Last event: {last_event}")

```

✓ 0.2s

```

First event: {'Apply for Viewing Appointment': 10000}
Last event: {'Reject Prospective Tenant': 6590, 'Tenant Cancels Apartment': 3273, 'Evict Tenant': 137}
First event: {'Apply for Viewing Appointment': 10000}
Last event: {'Reject Prospective Tenant': 5837, 'Tenant Cancels Apartment': 3942, 'Evict Tenant': 221}
First event: {'Apply for Viewing Appointment': 10000}
Last event: {'Reject Prospective Tenant': 7230, 'Tenant Cancels Apartment': 2680, 'Evict Tenant': 90}

```

Figure 8 : Liste des event de début et de fin par log

Une statistique un peut moins importante que nous pouvons obtenir est la répartition du nombre d'event par jour de la semaine:

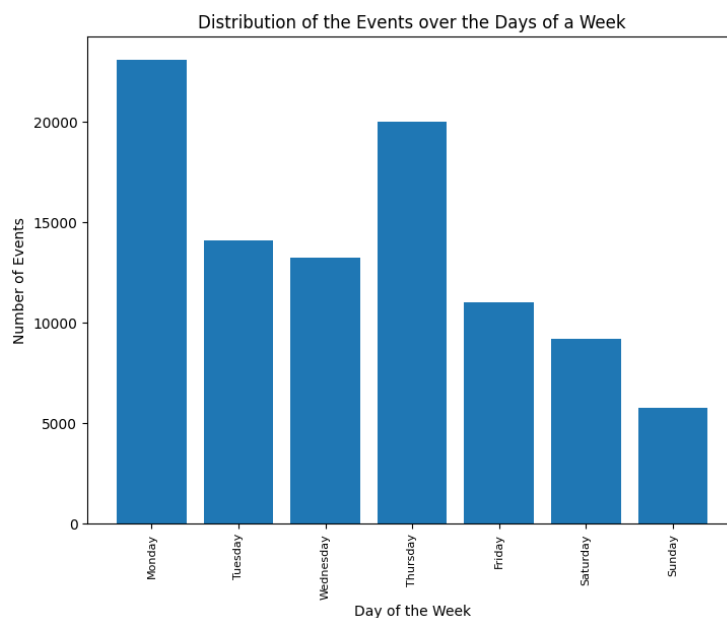


Figure 9 : Répartition par jour de la semaine



Il est aussi intéressant d'obtenir le tableau de footprint obtenu grâce à alpha miner. Nous pouvons ainsi y constater les relations entre les divers événements. Cette approche est cependant plus difficile à comprendre que le Pétri Net découvert plus tôt.

Python

	Accept Late Payment	Apply for Viewing Appointment	Check Paperwork	Exit Tenant	Extensive Screening	Hand In The Paperwork	Issue Warning	Move Into Payment	Move In/Pay Rent	Reject Prospective Tenant	Screen Prospective Tenant	Set Appointment	Sign Contract	Tenant Cancels Appointment	View The Property
Accept Late Payment	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apply for Viewing Appointment	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Check Paperwork	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Exit Tenant	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Extensive Screening	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hand In The Paperwork	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Issue Warning	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Move Into Payment	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Move In/Pay Rent	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Reject Prospective Tenant	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Screen Prospective Tenant	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Set Appointment	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sign Contract	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tenant Cancels Appointment	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
View The Property	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 10 : Footnote de renting low

Note: L'image est consultable sur le [GitHub](https://github.com) ou dans le rendu en meilleure qualité.

## Fuzzy Miner

Nous avons appliqué le fuzzy miner sur les données "high".

Voici la vidéo d'animation avec ProM (sur YouTube) :

<https://youtu.be/iOQA9-Wzx6k>

## Clustering

Nous avons appliqué le Clustering comme vu en cours et appliquons le coefficient de silhouette :

Silhouette coefficient pour 2 clusters: 0.6760900410843699

Silhouette coefficient pour 3 clusters : 0.6599244362280945

Une valeur proche de +1 indique que les clusters sont bien séparés

Pour 2 clusters nos clusters sont mieux séparés, nous allons donc utiliser ce nombre de clusters pour faire le dendrogramme.

Voici notre dendrogramme produit suivit du code du clustering :

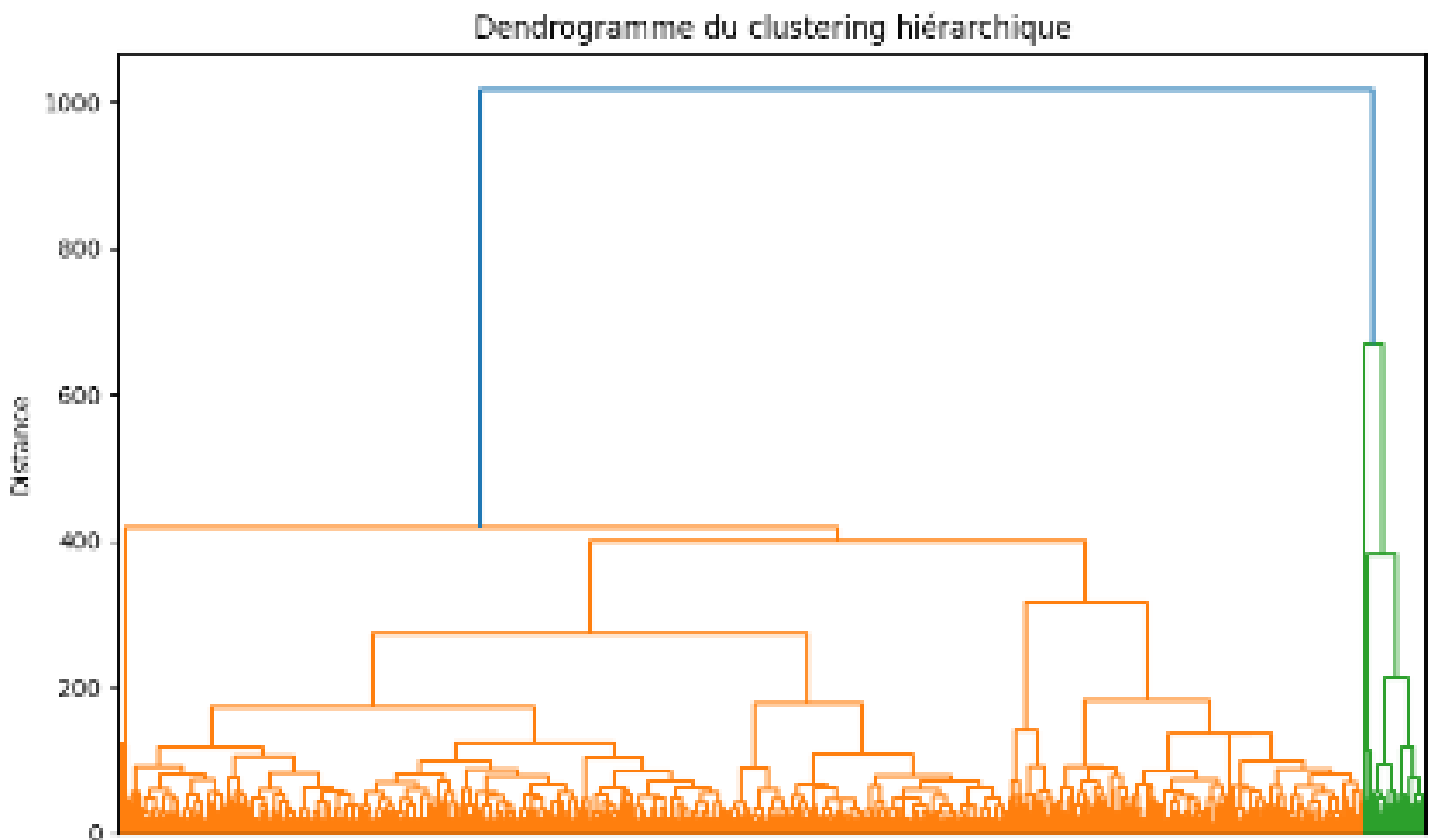


Figure 11 : Dendrogramme issu du clustering hiérarchique

```
#Clustering with decision point mining

from sklearn.cluster import AgglomerativeClustering

import pandas as pd

import numpy as np

df = log_high

variants = variants_filter.get_variants(df)

variants_dict = {v: i for i, v in enumerate(variants)}

variants_list = [v for v in variants]
```

```

# Define a function to compute Levenshtein distance
def levenshtein_distance(s1, s2):
    if len(s1) < len(s2):
        return levenshtein_distance(s2, s1)

    if len(s2) == 0:
        return len(s1)

    previous_row = range(len(s2) + 1)
    for i, c1 in enumerate(s1):
        current_row = [i + 1]
        for j, c2 in enumerate(s2):
            insertions = previous_row[j + 1] + 1
            deletions = current_row[j] + 1
            substitutions = previous_row[j] + (c1 != c2)
            current_row.append(min(insertions, deletions, substitutions))
        previous_row = current_row

    return previous_row[-1]

# Compute the Levenshtein distance matrix
distance_matrix = np.zeros((len(variants_dict), len(variants_dict)))
for i, s1 in enumerate(variants_dict):
    for j, s2 in enumerate(variants_dict):
        distance_matrix[i][j] = levenshtein_distance(s1, s2)

# Apply agglomerative clustering
num_clusters = 2

```

```
clustering = AgglomerativeClustering(n_clusters=num_clusters, linkage='complete')
cluster_assignments = clustering.fit_predict(distance_matrix)

# Silhouette coefficient
from sklearn.metrics import silhouette_score

print("Silhouette coefficient: {}".format(silhouette_score(distance_matrix,
cluster_assignments)))

#produire le dendrogramme
from scipy.cluster.hierarchy import dendrogram

import matplotlib.pyplot as plt

from scipy.cluster import hierarchy

# Calcul de la matrice de liaison
linkage_matrix = hierarchy.complete(distance_matrix)

# Affichage du dendrogramme
plt.figure(figsize=(10, 6))

dn = hierarchy.dendrogram(linkage_matrix, labels=variants_list, leaf_rotation=90)

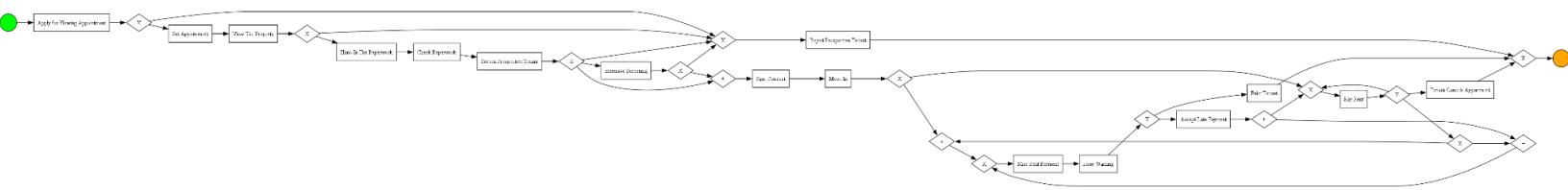
plt.title("Dendrogramme du clustering hiérarchique")

plt.xlabel('Variants')

plt.ylabel('Distance')

plt.show()
```

## Analyse sur le BPMN



BPMN généré avec l'Heuristic Miner, on remarque aucune anomalie spécifique dans le processus.

Process Flow:

Déroulement du processus :

1. Demande de rendez-vous pour la visite : Le processus commence par une demande de rendez-vous pour visiter une propriété.
2. Fixer le rendez-vous : Si la demande est acceptée, un rendez-vous est fixé.
3. Visiter la propriété : Le locataire potentiel visite la propriété.
4. Remettre les documents : Après la visite, le locataire potentiel remet les documents nécessaires.
5. Sélectionner le locataire potentiel / Vérification pour l'entreprise : Les documents entraînent un processus de vérification, qui semble avoir deux branches - une pour le locataire et une pour l'aspect commercial.
6. Signature du contrat : Si les vérifications sont concluantes, le contrat est signé.
7. Emménager : Le locataire emménage dans la propriété.
8. Accepter un paiement tardif / Émettre un avertissement : En cas de retard de paiement, il y a une décision à prendre pour accepter le paiement tardif ou émettre un avertissement.
9. Expulser le locataire / Payer le loyer : En cas de problèmes persistants, il y a un point de décision pour l'expulsion ou le paiement du loyer par le locataire.
10. Le locataire annule l'appartement : Le processus se termine si le locataire annule l'appartement.

Problèmes potentiels ou violations de schémas :

- Prise de décision complexe : Il y a plusieurs points de décision qui pourraient devenir complexes, surtout qu'il n'y a pas de directives claires sur quand accepter les paiements tardifs ou émettre des avertissements.
- Bouclage : Le processus semble avoir plusieurs boucles potentielles, notamment autour de la section des paiements.
- Conditions de fin : La fin du processus n'est pas clairement définie pour une location réussie, seulement pour l'annulation.
- Logique de ramification : Il y a plusieurs points où le processus se divise, et il n'est pas tout à fait clair comment chacun devrait être traité pour assurer un flux logique.

## Analyse des discrimination

Afin de se représenter nos divers résultats nous avons créé des fonctions pour print le nombre de cas en fonction de certains critère ainsi que les pourcentage d'occurrence dans l'event log.

Dans ces analyses nous reconnaissons les events suivant comme indicateurs de discrimination en fonction de leur nombre d'occurence et de la différence de valeur obtenue.

Nom de l'event	Signification
Sign Contract	Le contrat à été signé suite à un accord
Reject Prospective Tenant	À n'importe quel moment du processus le propriétaire refuse la location
Evict Tenant	Le locataire est expulsé (souvent à cause d'impayés)

Issue Warning	Le propriétaire donne un avertissement d'impayé du loyer.
---------------	---

Ces événements appartiennent à 2 catégories de critères.

- L'étude d'acceptation (en bleu ci dessus) étudie le biais à accepter et refuser la signature
- L'étude d'expulsion (en vert) étudie la tendance d'expulsion.

Nos résultats ont d'abord été rendus sur un notebook mais nous les avons ensuite ajoutés dans un classeur Google Sheets. Afin notamment de mettre en forme de représenter les résultats sous forme de graphiques.

## Statistiques générales

Par exemple, voici des stats générales disponibles [ici](#) dans l'onglet 'Stats'. Nos résultats sont ici représentés en pourcentage.

Dans l'exemple ci-contre nous constatons par exemple que le nombre de warning et d'expulsion par case sont plutôt similaires. En outre nous discernons qu'en moyenne un locataire a un peu plus d'un retard de paiement sur tout le processus de location.

### Stats générales

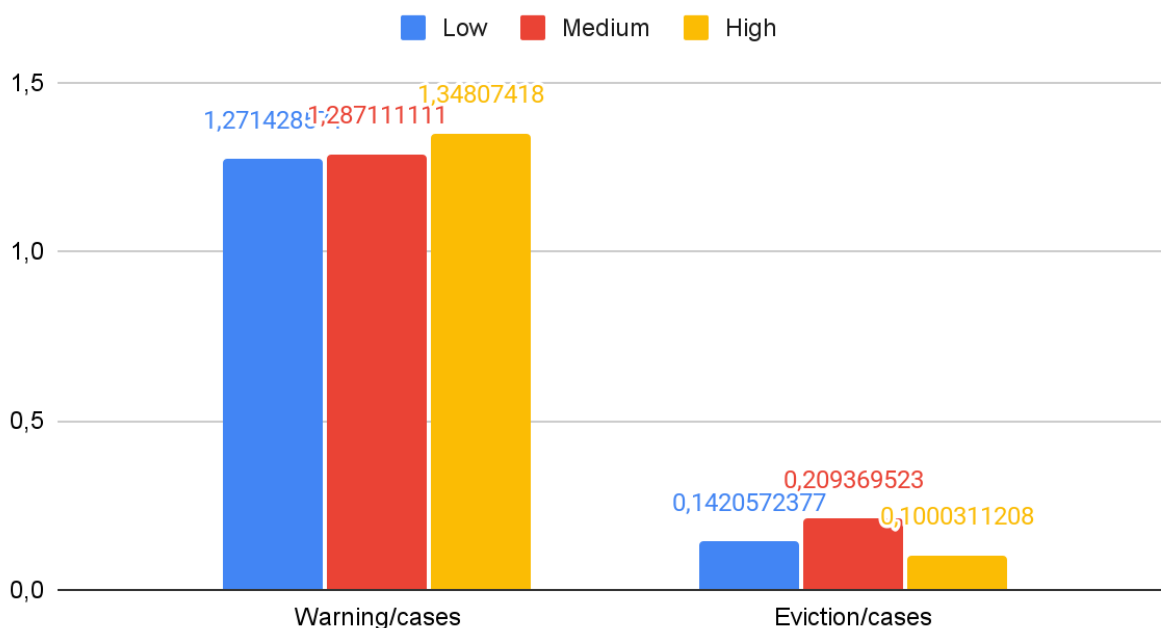


Figure 12 : Pourcentage de cas avec expulsion ou avertissement pour chaque dataset.

En plus de ces résultats nous avons une feuille entièrement dédiée aux statistiques générales sous forme tabulaire.

	Low	Medium	High
<b>Nb cases</b>	10000	10000	10000
<b>Nb events</b>	96440	105555	610
<b>Nb variants</b>	508	89972	496
Warning/cases	1,271428571	1,287111111	1,34807418
Eviction/cases	0,1420572377	0,209369523	0,1000311208



## Sexisme

Comme mentionné plus tôt l'étude des discriminations se fait sur le notebook, dans la capture ci dessous nous affichons les résultats d'une fonction mettant en évidence la différence de traitement entre les homme (dénuté ♂) et les femmes (dénuté ♀).

```

gender
Dataset: low
Value: False
- Accepted: 2523 → 56.40500000000001 %
- Denied: 1950 → 43.595 %

Value: True
- Accepted: 3197 → 57.843 %
- Denied: 2330 → 42.157 %

Dataset: medium
Value: False
- Accepted: 3370 → 65.988 %
- Denied: 1737 → 34.012 %

Value: True
- Accepted: 3209 → 65.583 %
- Denied: 1684 → 34.416999999999994 %

Dataset: high
Value: False
- Accepted: 2185 → 51.888 %
- Denied: 2026 → 48.112 %

Value: True
- Accepted: 3004 → 51.892 %
- Denied: 2785 → 48.108000000000004 %

gender
Dataset: low
Value: False
- Evicted: 78 → 3.092 %
- Canceled: 2445 → 96.908 %

Value: True
- Evicted: 125 → 3.91 %
- Canceled: 3072 → 96.09 %

Dataset: medium
Value: False
- Evicted: 190 → 5.638 %
- Canceled: 3180 → 94.362 %

Value: True
- Evicted: 200 → 6.232 %
- Canceled: 3009 → 93.768 %

Dataset: high
Value: False
- Evicted: 86 → 3.936 %
- Canceled: 2099 → 96.06400000000001 %

Value: True
- Evicted: 91 → 3.029 %
- Canceled: 2913 → 96.97099999999999 %

```

Ce format est lourd et difficile à lire. Nous l'envoyons dans le classeur et nous en faisons un graphe.

Voici un exemple des données au format tabulaire.

	Low	Medium	High
<b>Nb cases</b>	10000	10000	10000
<b>Nb events</b>	96440	105555	610
<b>Nb variants</b>	508	89972	496
<b>Nb acceptance</b> ♂	2523	3370	2185
<b>Nb refus</b> ♂	1950	1737	2026
<b>Total Traces</b> ♂	4473	5107	4211

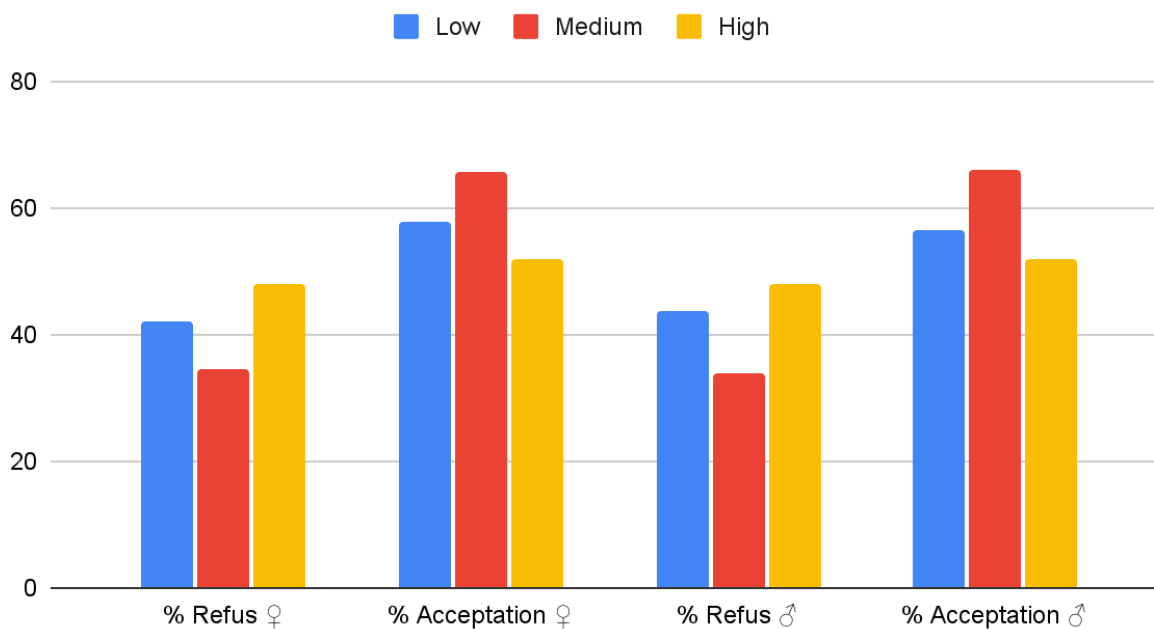
<b>Nb acceptation</b> ♀	3197	3209	3004
<b>Nb refus</b> ♀	2330	1684	2785
<b>Total Traces</b> ♀	5527	4893	5789
<b>% Acceptation</b> ♂	56,40509725	65,9878598	51,88791261
<b>% Refus</b> ♂	43,59490275	34,0121402	48,11208739
<b>% Acceptation</b> ♀	57,84331464	65,58348661	51,8915184
<b>% Refus</b> ♀	42,15668536	34,41651339	48,1084816

Ce qui nous donne ce graphe.

Nous n'y constatons que des écart très faible, en moyenne dans low les femmes se font moins souvent refuser la signature et les hommes signent moins de contrats que les femmes.

Les résultats dans les autres graphes (medium et high) ne montre aucune autre discrimination.

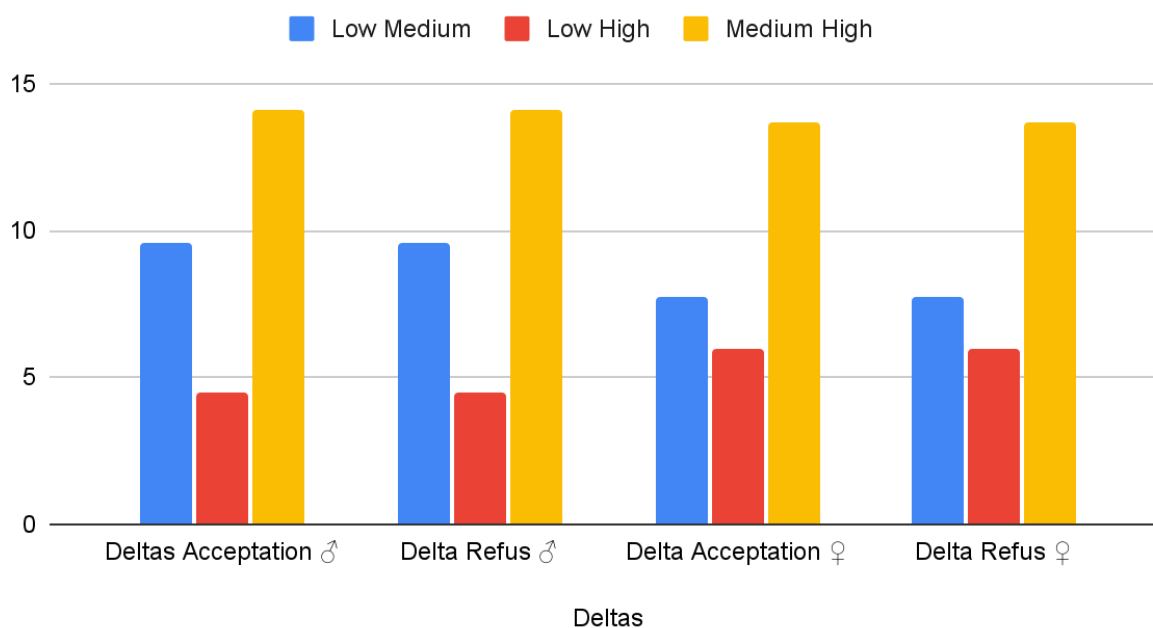
## Taux d'acceptation et de refus des Hommes et des Femmes



Un autre résultat intéressant est de comparer la différence entre les différents logs.

Dans le graphe ci-dessous nous constatons que la plus grande différence de traitement entre les hommes et les femmes n'est pas entre le log 'Low' et 'High', mais entre 'High' et 'Medium'. Ce qui est contre intuitif et remet en question le choix du nom du fichier.

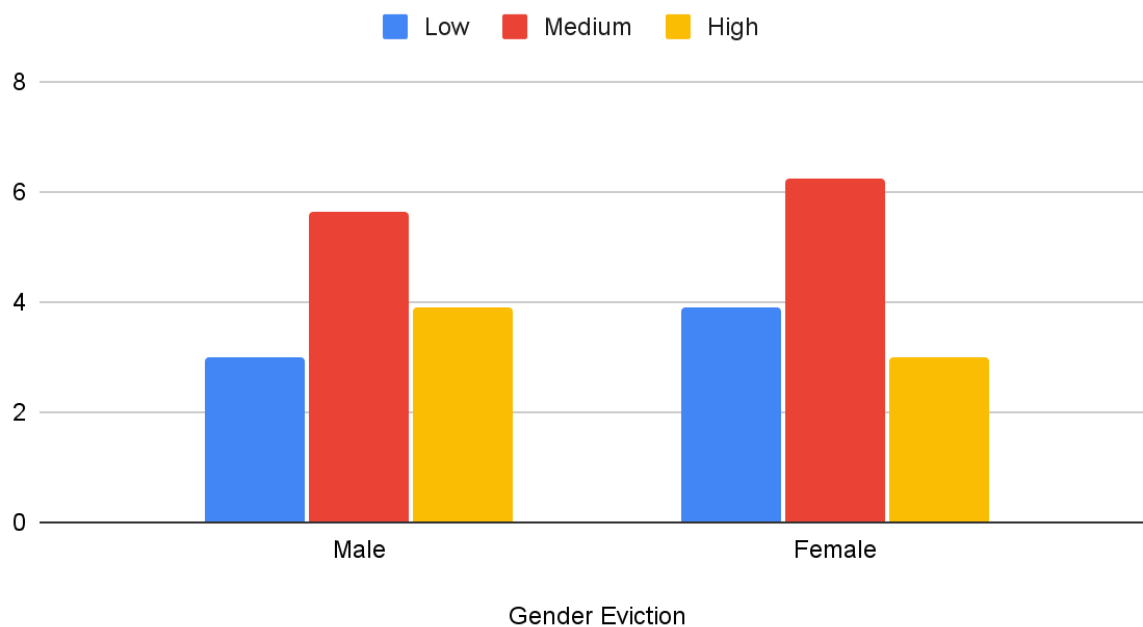
### Analyses des différences de % entres les différents logs



Nous pouvons également tenter de faire une analyse sur les exclusions par genre.

Gender Eviction	Low	Medium	High
Male	3	5,64	3,9
Female	3,91	6,23	3

## Taux d'eviction en % en fonction du genre



Ce critère met en évidence de légers biais, dans low et medium, les hommes sont moins souvent expulsés que les femmes. Mais dans High c'est l'inverse.

Ces graphes et valeurs sont disponibles dans l'onglet 'Gender' du classeur.

## Âgisme

Nous proposons également une étude de la discrimination par tranche d'âge.

D'abords nous etudions la discrimination en fonction de l'acceptation et du refus de signature de contrat avec des tranches d'âges définies tel que:

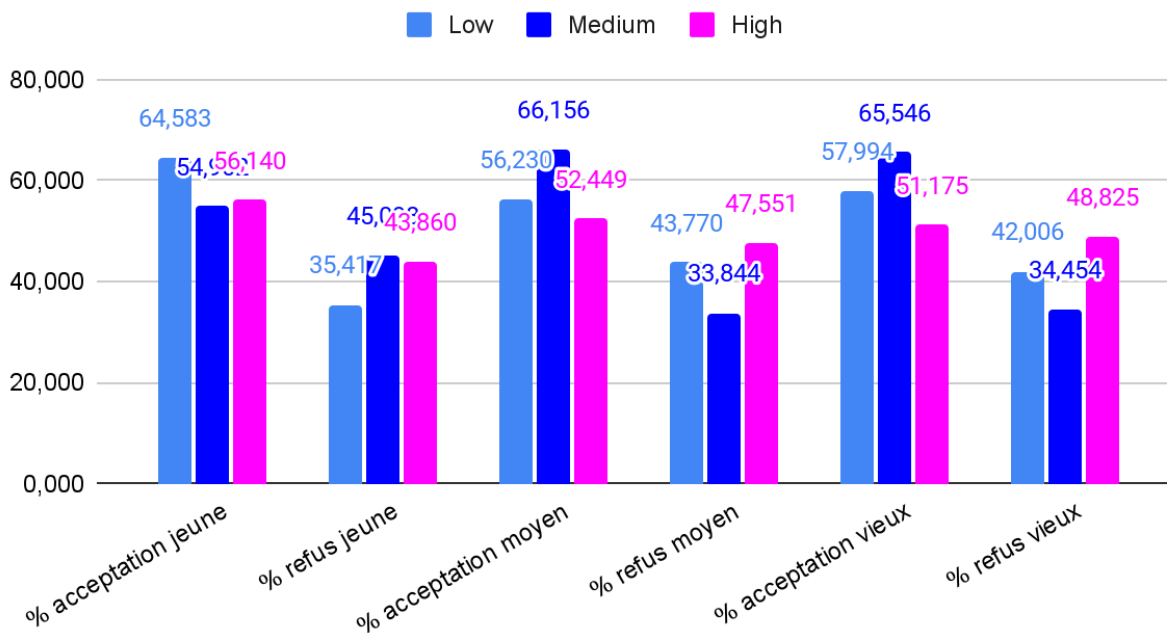
- Tranche d'âge < 25 ans :  $x \in [0, 25[$
- Tranche d'âge > 25 ans et < 50 ans :  $x \in ]25, 50[$
- Tranche d'âge > 50 ans :  $x \in ]50, +\infty[$

D'après nos recherches, les discriminations ne sont pas évidentes selon ce critère.

Nous constatons dans low que les jeunes signent plus fréquemment. Mais l'échantillon n'est pas très représentatif à cause du positionnement de l'intervalle. Les jeunes ne sont donc pas très représentatifs. Hormis ces éléments il n'y a pas de discrimintaion évidente.

	Low	Medium	High
Acceptation jeune	31	28	32
Refus jeune	17	23	25
Total jeune	48	51	57
Acceptation moyen	2631	3237	2827
Refus moyen	2048	1656	2563
Total moyen	4679	4893	5390
Acceptation vieux	3058	3314	2330
Refus vieux	2215	1742	2223
Total vieux	5273	5056	4553
% acceptation jeune	64,583	54,902	56,140
% refus jeune	35,417	45,098	43,860
% acceptation moyen	56,230	66,156	52,449
% refus moyen	43,770	33,844	47,551
% acceptation vieux	57,994	65,546	51,175
% refus vieux	42,006	34,454	48,825

## Taux d'acceptation et de refus en fonction des tranches d'âges



Les résultats n'étant peut concluants nous avons décidé de faire une analyse du taux d'expulsion avec des tranches d'âges plus minces.

Voici les tranches d'âge définies dans le code sous forme de texte avec des formules :

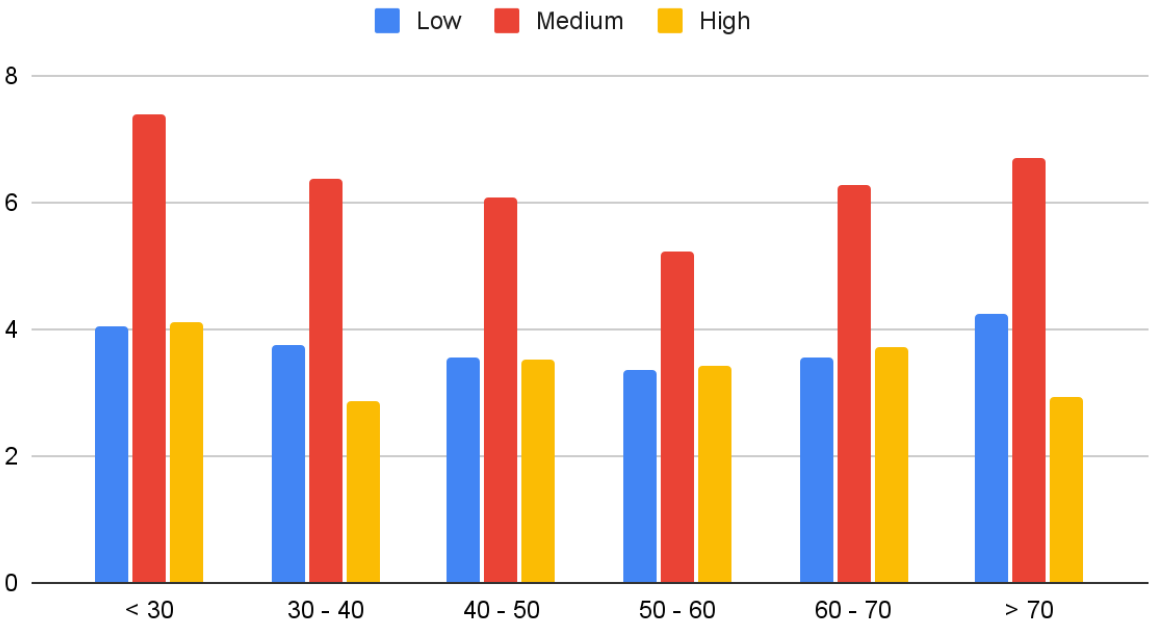
Nom	Tranche basse	Tranche haute	Formule
<b>Jeune adulte</b>	20	30	age_jeune=age[(age≥20) & (age≤30)]
<b>Adulte d'âge moyen</b>	30	40	age_moyen=age[(age≥30) & (age≤40)]
<b>Personne âgée</b>	40	50	age_vieux=age[(age≥40) & (age≤50)]
<b>Personne très âgée</b>	50	60	age_tres_vieux=age[(age≥50) & (age≤60)]
<b>Ancien</b>	60	70	age_ancien=age[(age≥60) & (age≤70)]
<b>Très ancien</b>	70	80	age_tres_ancien=age[(age≥70) & (age≤80)]

Note: Nous nous dédouanons de tout dégâts émotionnels engendré par la dénomination susmentionné

Les résultats sur le taux d'expulsions sont comme tel:

- Sur tous les dataset, on observe une diminution des expulsions pour les locataires d'âge moyen ou âgés (30-50) en opposition aux jeunes adultes et personnes âgées.
- Sur le dataset low, le taux d'expulsion est plutôt stable autour de 3.5%
- Sur le dataset medium, le taux moyen est plus élevé et les variations sont les plus marquées.

### Taux d'eviction en % en fonction de l'âge





## Discrimination par langue parlée

Nous nous intéressons dans cette partie à l'étude de potentiels biais envers les germanophones.

Nous définissons les termes suivants:

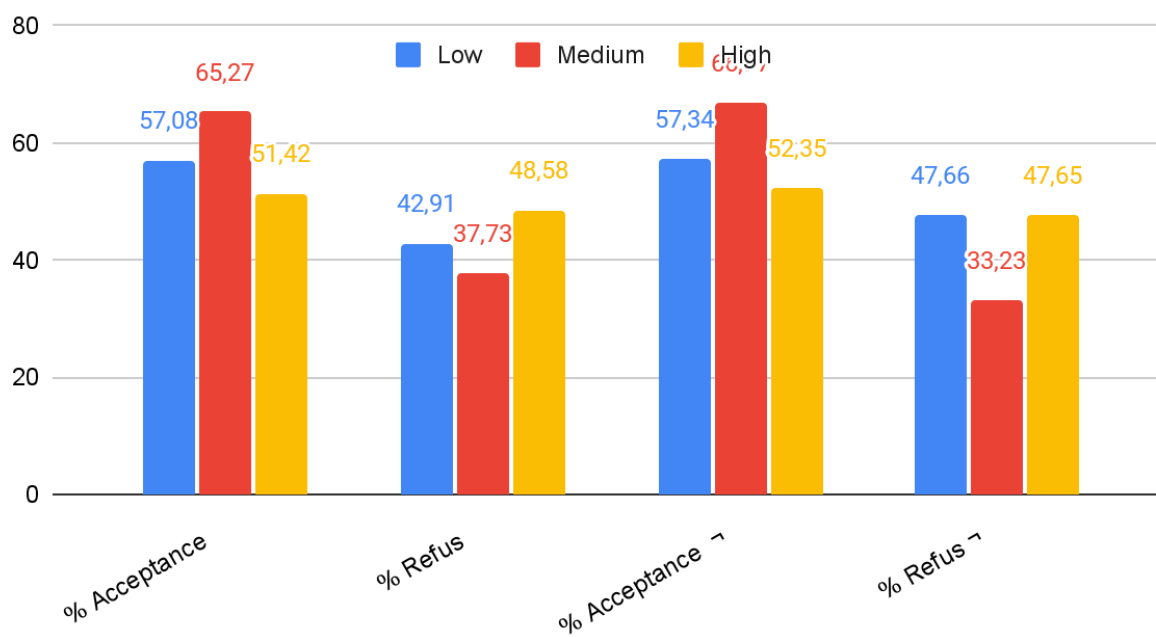
- Parle allemand : 🇩🇪, german speaking = True
- Ne parle pas allemand : ¬🇩🇪, german speaking = False

Nous faisons d'abord une étude en fonction du taux d'acceptation et refus de signature des contrats.

	Low	Medium	High
% Acceptance 🇩🇪	57,08	65,27	51,42
% Refus 🇩🇪	42,91	37,73	48,58
% Acceptance ¬🇩🇪	57,34	66,77	52,35
% Refus ¬🇩🇪	47,66	33,23	47,65



Nous obtenons ce graphe

### Discrimination germanophone en %

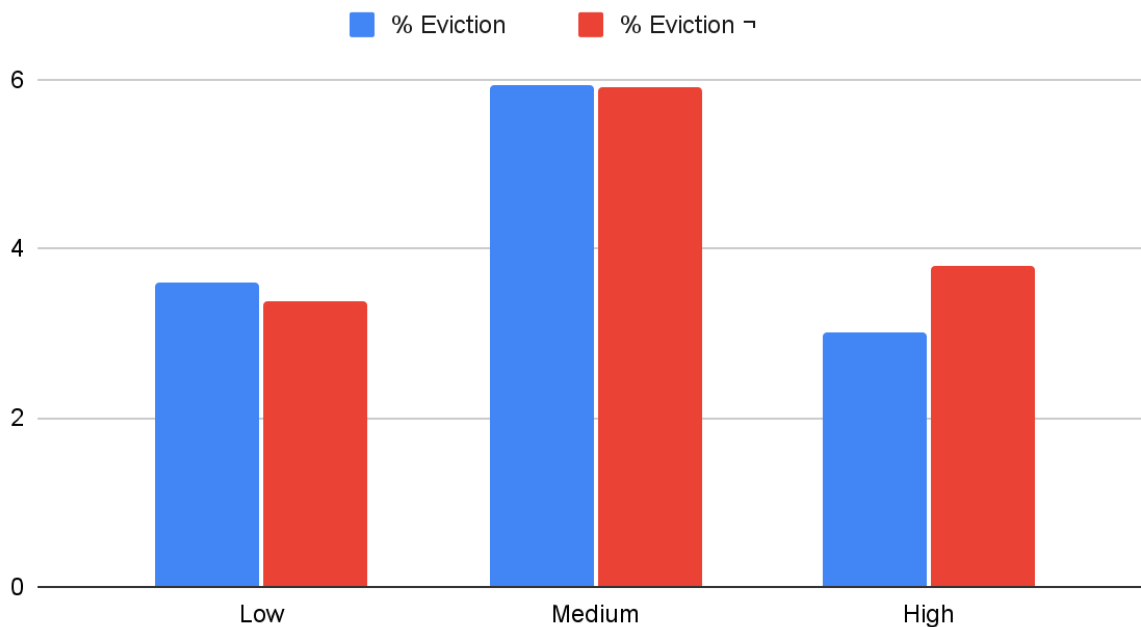


Les résultats sont mitigés, nous constatons dans le log médium un léger biais les refus pour les non germanophones sont moins fréquents. Mais dans le log low nous constatons l'inverse. En revanche, les taux d'acceptation sont entièrement homogènes entre les deux catégories.

Nous pouvons essayer de trouver des discriminations sur les taux d'expulsions avec les données ci-dessous:

	Low	Medium	High
% Eviction 	3,6	5,94	3,01
% Eviction 	3,37	5,91	3,79

### % Eviction et % Eviction



Nous concluons que les biais sont:

- Nul dans les log low
- Nul dans le log medium
- Léger dans le log high

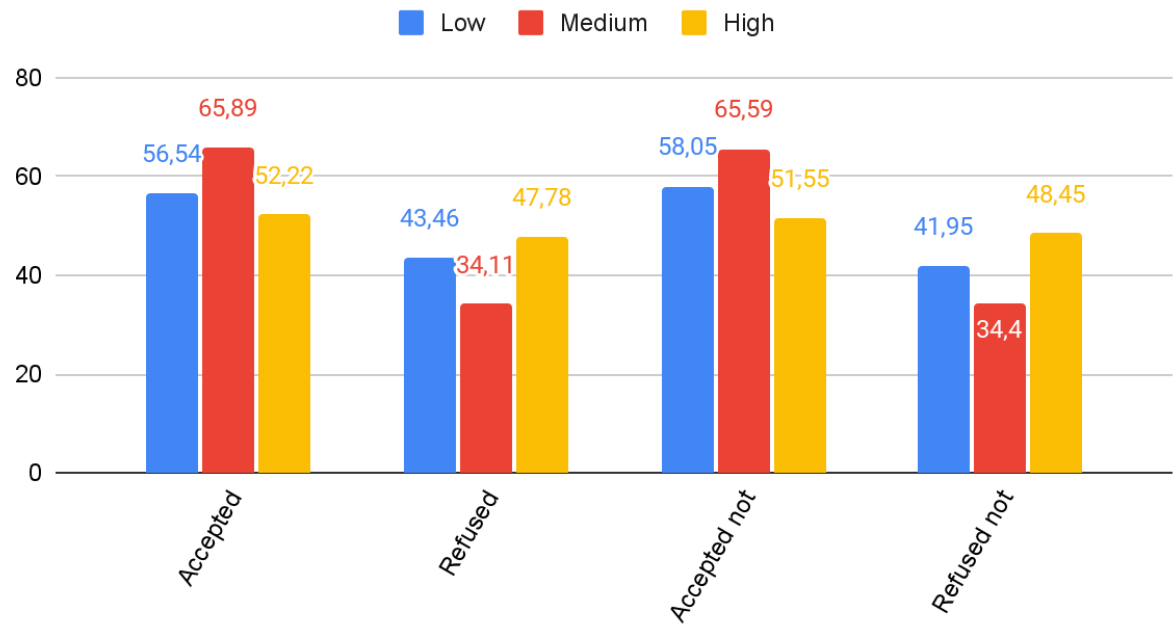
Discrimination par status marital

Le statut marital apparaît dans les logs, nous pouvons essayer d'étudier la tendance de signature.

	Low	Medium	High
Accepted married	56,54	65,89	52,22
Refused married	43,46	34,11	47,78
Accepted not married	58,05	65,59	51,55
Refused not married	41,95	34,4	48,45

Nous trouvons des résultats très similaires et donc peu concluants. Nous constatons que les résultats provenant de high sont bien plus homogènes à l'opposé de médium. Nous ne constatons cependant pas vraiment de biais selon ces critères.

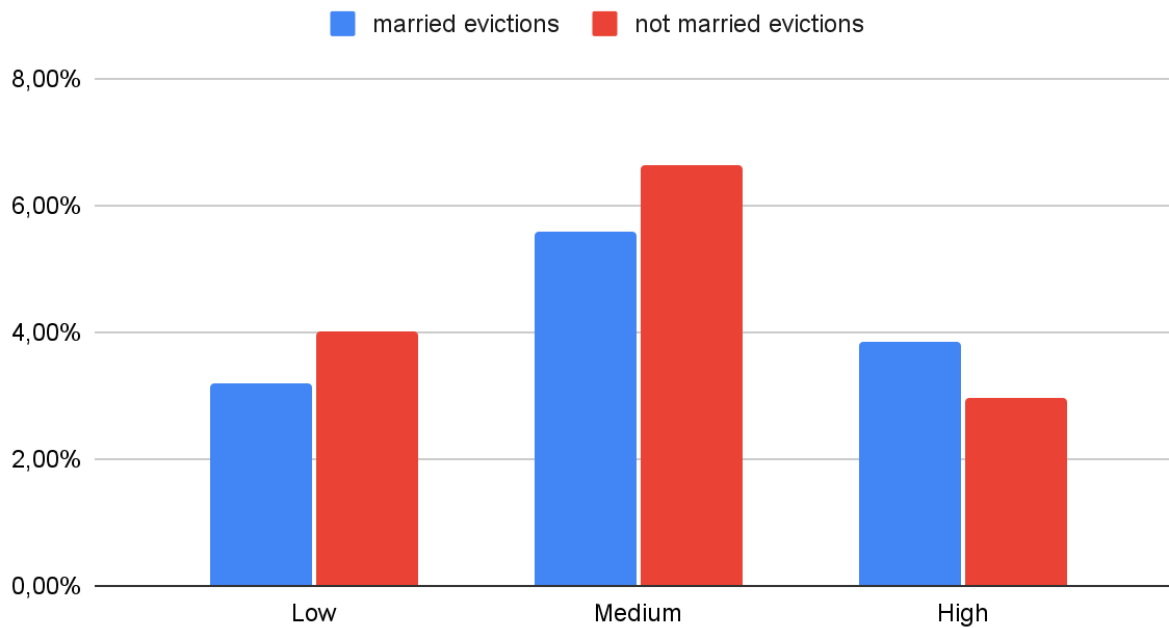
Acceptation par status marital



Nous faisons maintenant l'étude par expulsion.

	Low	Medium	High
<b>married evictions</b>	3,19%	5,57%	3,84%
<b>not married evictions</b>	4,00%	6,63%	2,97%

## Expulsions par statut marital



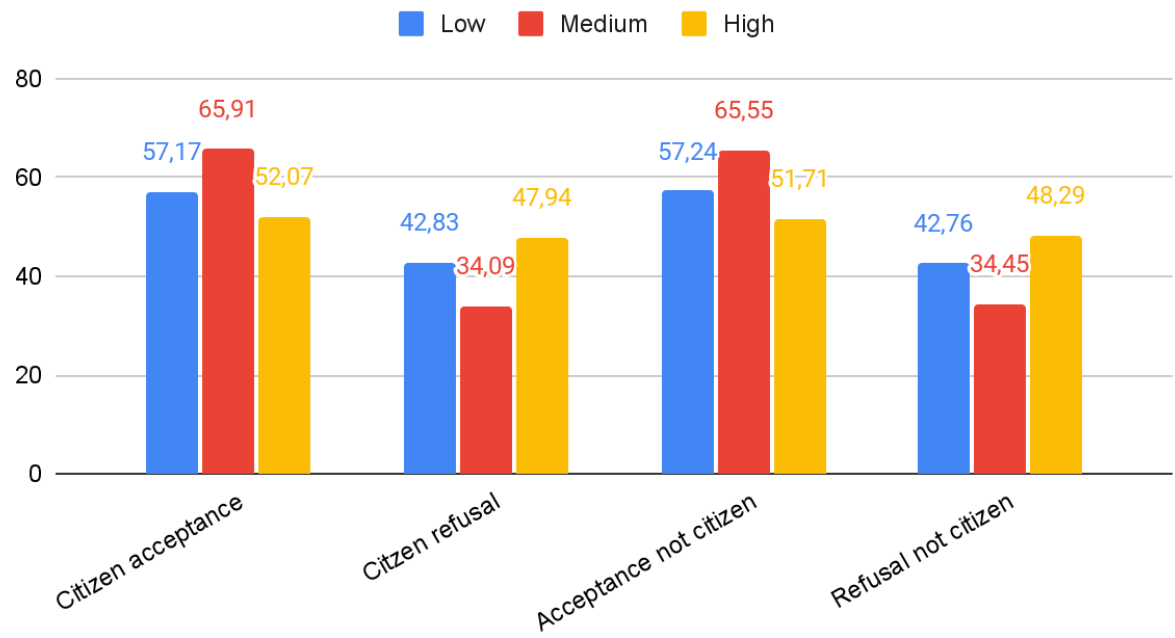
Discrimination par citoyenneté

Nous étudions la discrimination par citoyenneté. Définie comme une variable binaire au sein de nos logs.

	Low	Medium	High
Citizen acceptance	57,17	65,91	52,07
Citizen refusal	42,83	34,09	47,94
Acceptance not citizen	57,24	65,55	51,71
Refusal not citizen	42,76	34,45	48,29

Les résultats sont similaire aux deux dernières parties et ne nous permettent pas de conclure sur une forme de discrimination par citoyenneté.

Acceptation par citoyenneté

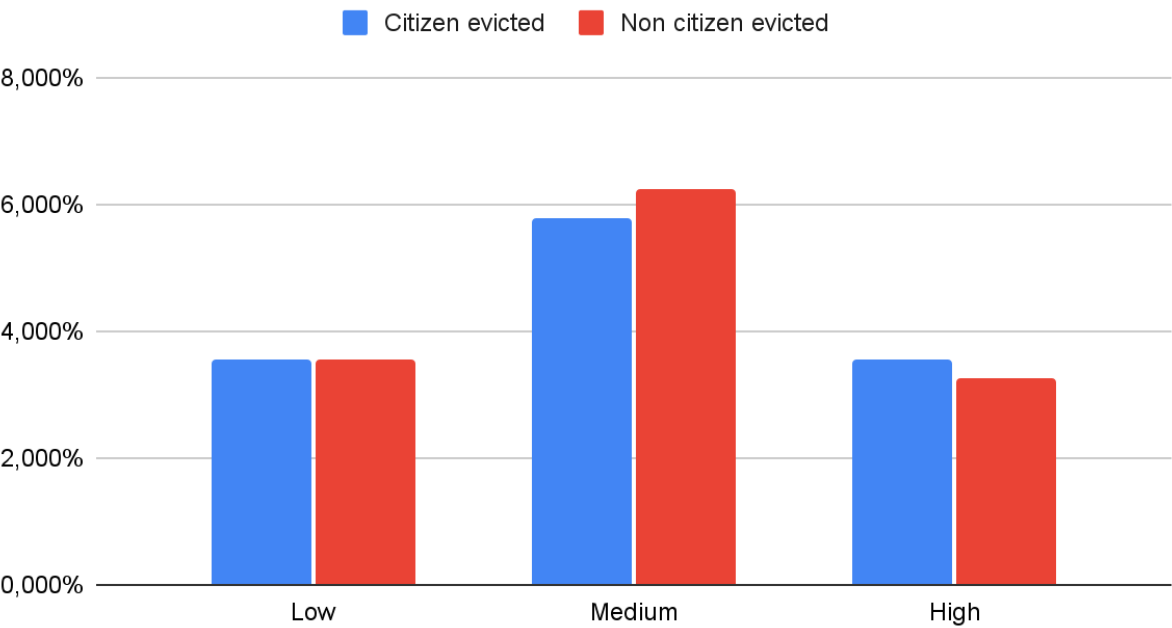


Nous étudions maintenant les expulsions par citoyenneté.

	Low	Medium	High
--	-----	--------	------

Citizen evicted	3,540%	5,770%	3,560%
Non citizen evicted	3,560%	6,240%	3,250%

Pourcentage d'expulsion par citoyenneté en %

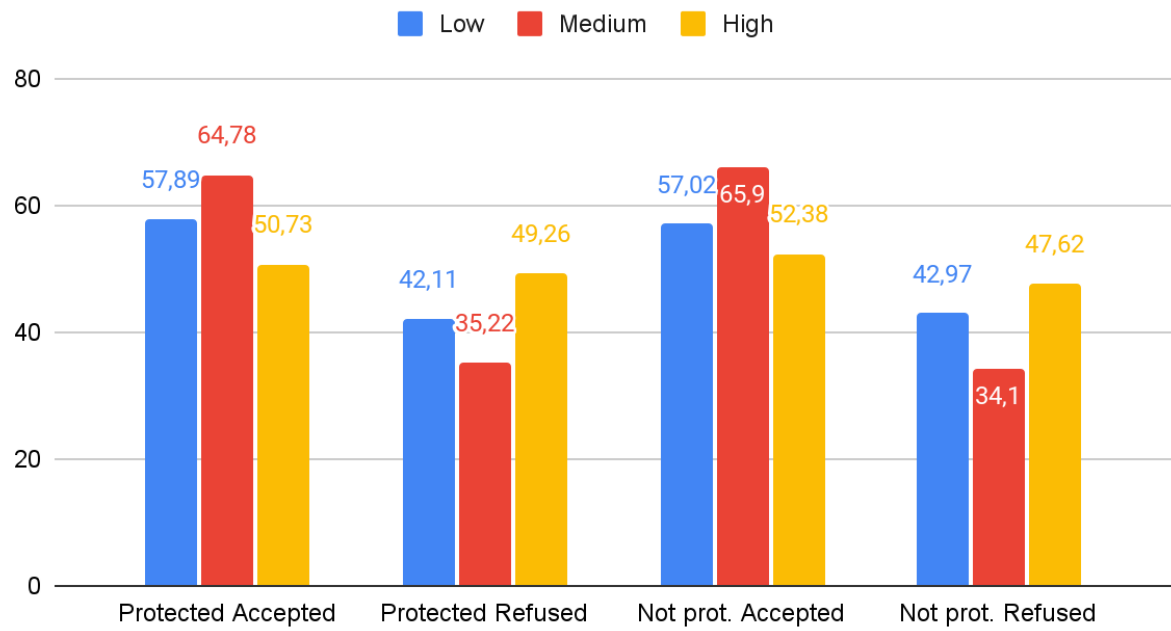


Les résultats ne mettent en avant aucune discrimination très significative. Nous voyons un très léger biais dans le log medium.

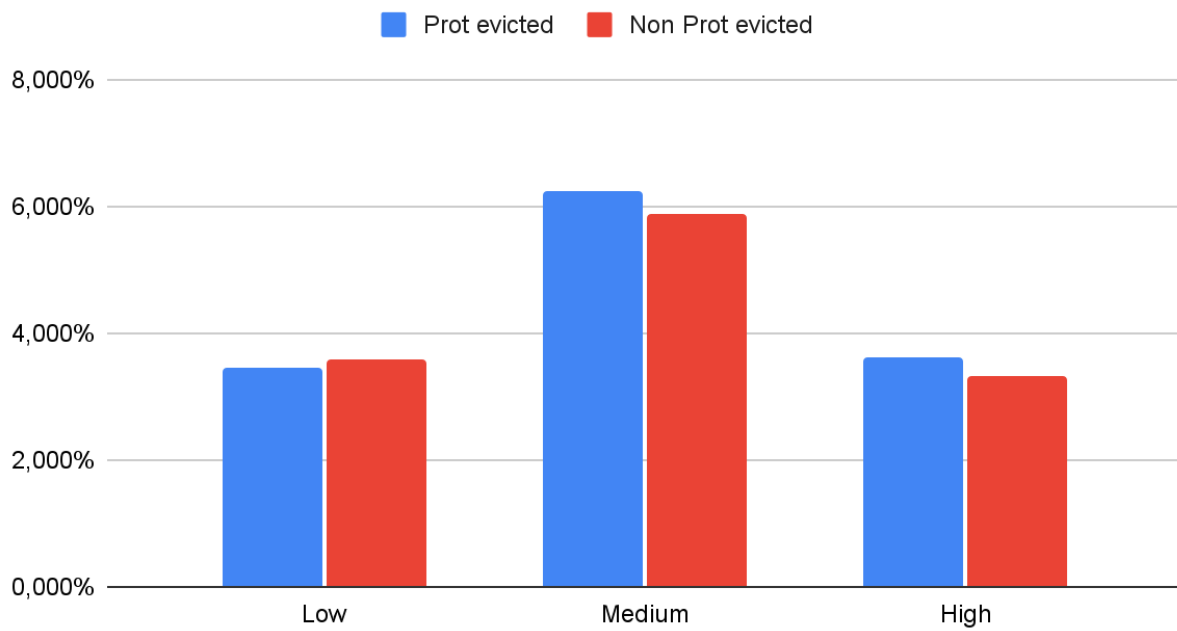
Discrimination par "protected"

Nous ne comprenons pas ce que ce paramètre.

Taux acceptation protected



## Taux eviction protected



Protected ne semble pas être un discriminant évident. Il existe cependant un très léger écart sur le taux d'acceptation dans High.



## Discrimination par années d'études

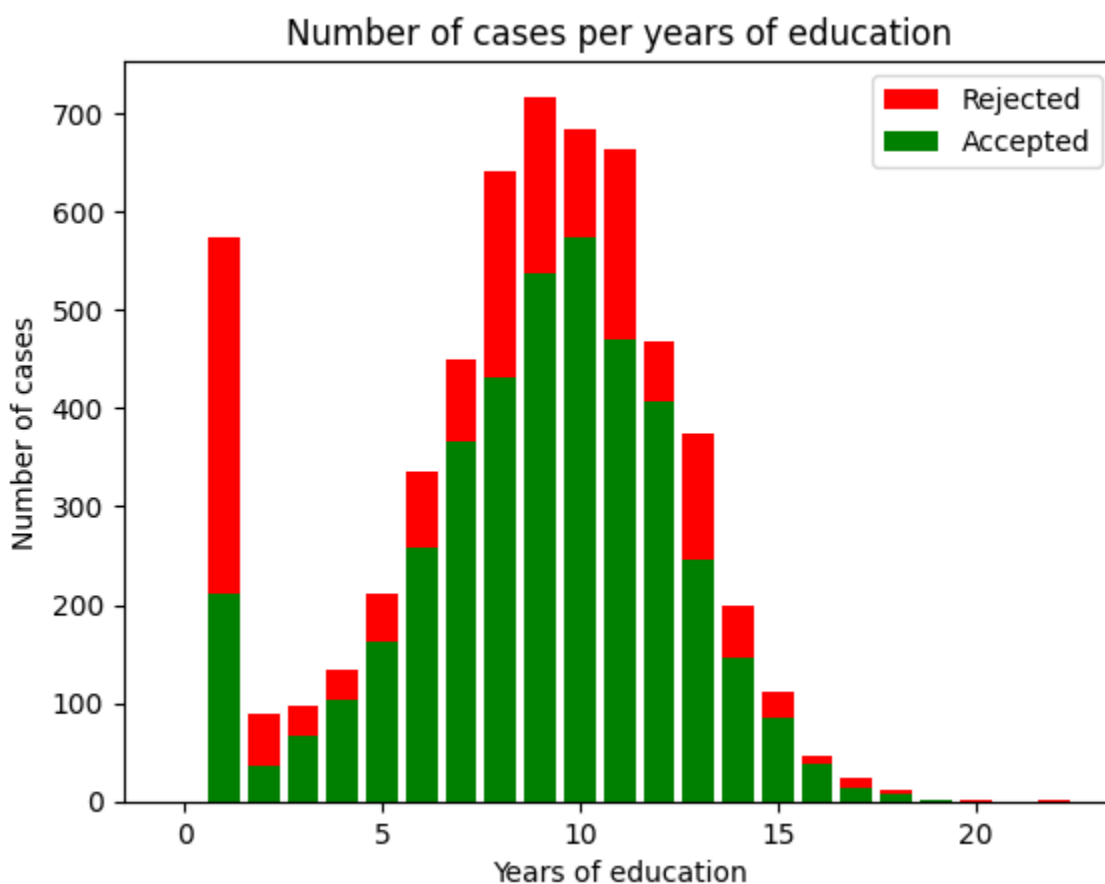
Les données se trouvent dans le fichier Excel dans la feuille Études.

Grâce à ces données, on peut remarquer que plus les années d'études augmentent et plus le taux d'acceptation évolue, bien que les données ne soient pas assez représentatives pour les très haut niveau d'études.

Cependant, on remarque aussi que via ces données, on peut voir que pour les très bas niveaux d'études les demandes sont plus acceptées que les niveaux moyens d'études. Nous soupçonnons un manque de précision dans la quantité des données.

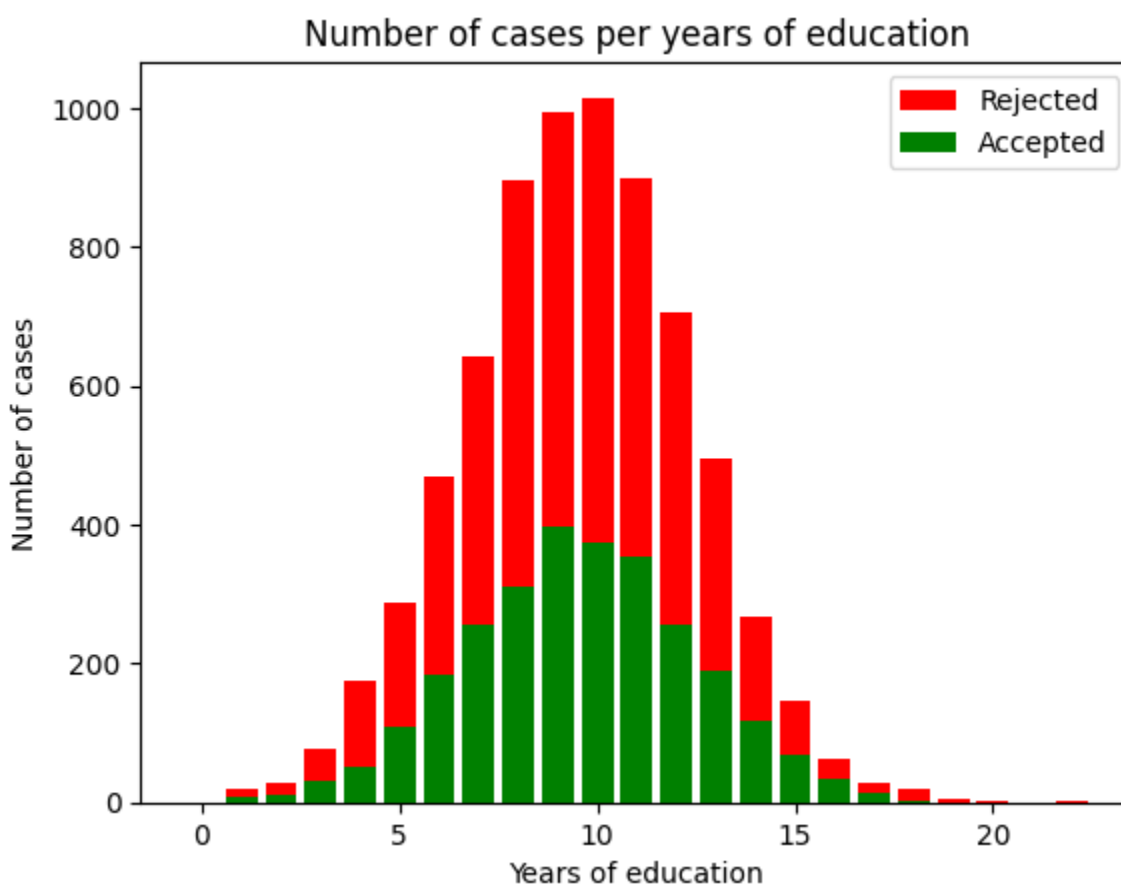
Dans le log "medium" il y a une grosses différences dans les données, avec un pic de demandes pour des niveaux très bas d'études.

Graphique log "medium" :

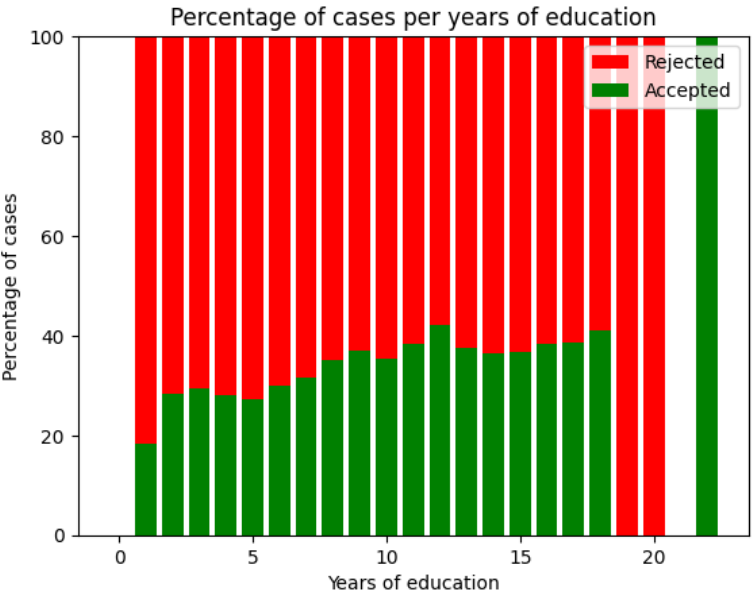


Cependant pour les données medium, nous trouvons un graphique “normé”, c'est-à-dire où nous retrouvons une norme correspondante aux nombres de demandes.

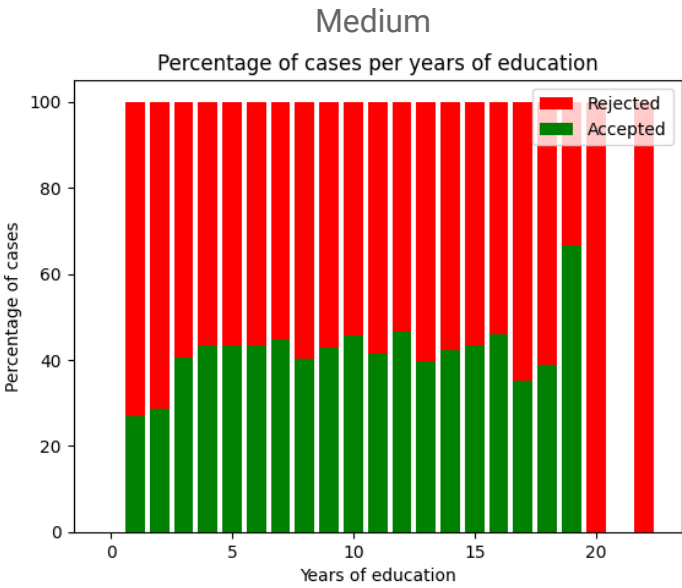
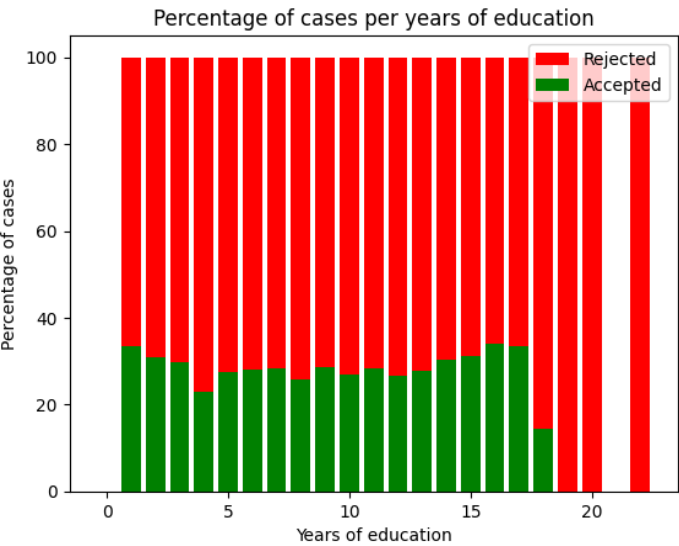
Bien que toujours avec la légère différences visibles dans les données sur le pourcentage d'acceptation avec des plus haut niveaux d'études :



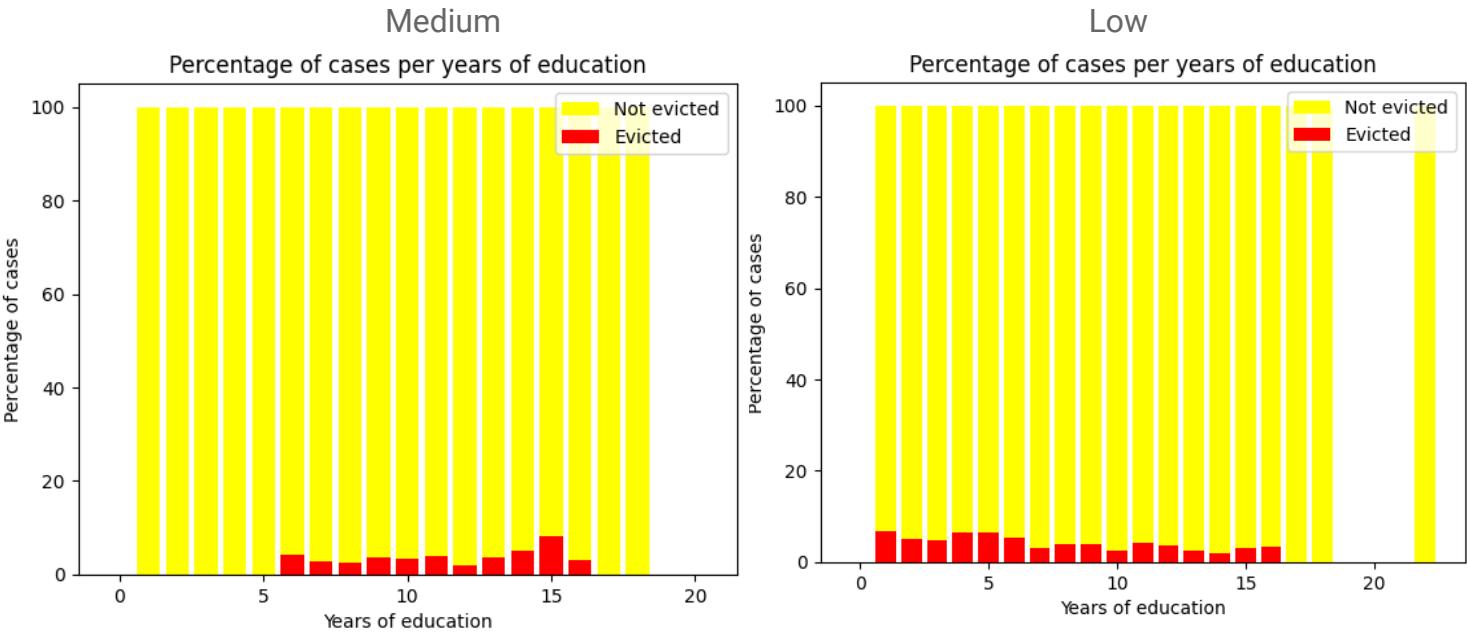
Avec ces 3 graphiques suivant, nous remarquons bien les différences entre les différents logs et on remarque aussi les pourcentages qui diffèrent selon les années d'études :



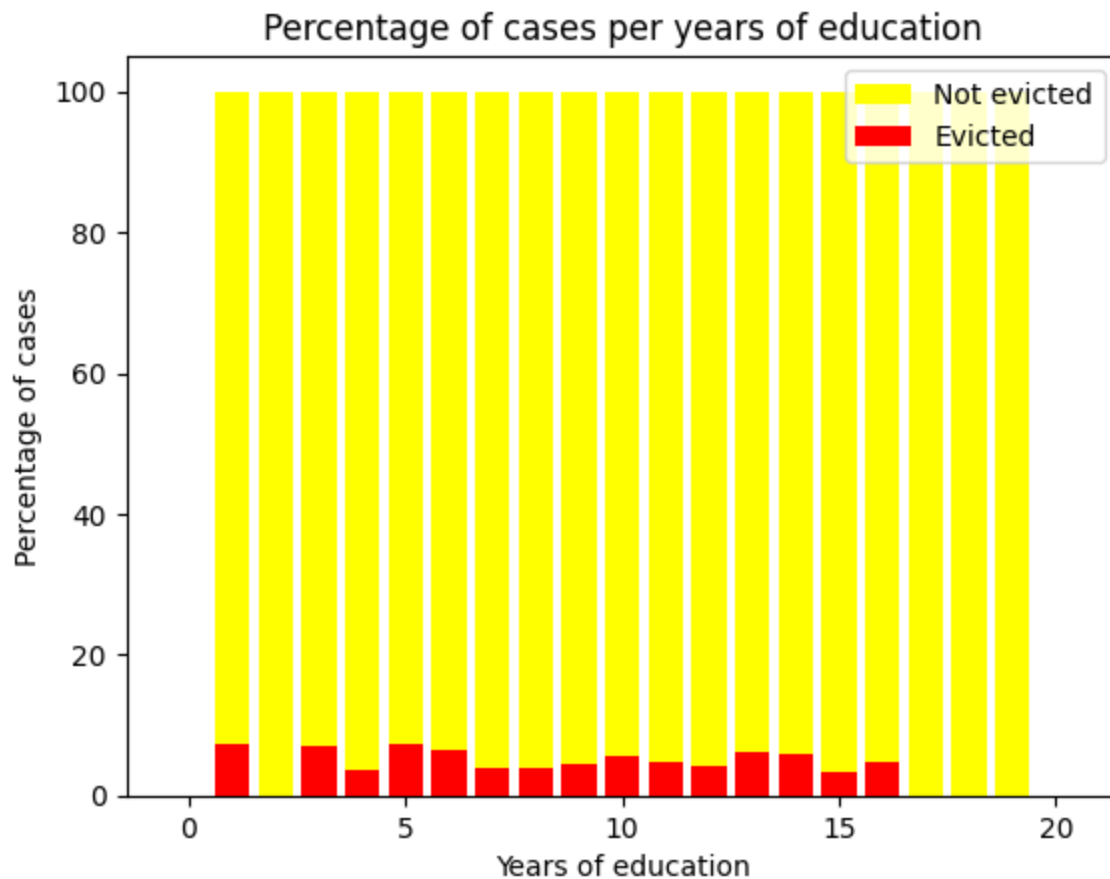
<- Low



Eviction



High



Selon les données "low", "medium", "high" on a différentes différenciations.

De ce fait, nous ne pouvons pas conclure quelque chose spécifiquement.

Cependant, pour "low", on remarque que plus, on a d'années d'études, moins on a de risque d'être dans l'état "evicted".

Néanmoins, on a aussi l'inverse sur le "medium" et pour "high" ça paraît aléatoire...

On peut en conclure qu'il y a un manque de données pour représenter les différents cas.

## Récapitulatif

Ce tableau fait la synthèse des résultats obtenus vis-a-vis des différents paramètres des traces.

Discrimination ?	Étude d'acceptation			Étude des expulsions		
Log / Critère	Low	Medium	High	Low	Medium	High
Genre	Faible	Non	Non	Faible	Faible	Faible
Âge	N/R	Non	Non	T Faible	Faible	Non
Language	Non	Non	T Faible	Non	Non	Faible
Mariée	Non	Non	Non	Faible	Faible	Faible
Citoyen	Non	Non	Non	Non	T Faible	Non
Protected	Non	Non	T Faible	Non	Non	Non
Études	T Faible	Notable	Faible	Notable	T Faible	Non
Scores	3	3	4	8	8	6

N/R = Échantillons trop peu nombreux

## Conclusion

D'après nos résultats précédents l'évènement définissant le plus les discrimination est le taux d'expulsion. Le nombre d'années d'études, le genre et le statut marital semblent êtres les facteurs de discrimination les plus importants.

Le taux d'acceptation et de refus met en évidence moins de discrimination, sauf dans l'étude de la discrimination par niveau d'étude. Dans ce cas nous voyons une discrimination plutôt évidente.