

ANALYSE EXPLORATOIRE DES DONNÉES

Devoir 01

Étudiant:

Mohamed Toros

Professeur

Pierre-Marc Juneau

01 Avril 2021

TABLE DES MATIÈRES

Préparation et prétraitement des données	3
Nettoyage :	3
Exploration :	3
Variables et facteurs influençant la productivité	3
Analyse de sensibilité (Monté Carlo) (code script python)	6
Impact du prix du bois (\$ / pmp) sur la distribution des profits	6
Note technique	8

Préparation et prétraitement des données

A ce niveau on a procédé à la consolidation des données, des trois fichiers (fournisseur A, B et C) qu'on la même structure

Nettoyage :

A ce niveau on a procédé, au nettoyage des valeurs manquantes (stratégies d'enlever les tuples avec Nan).
Décision prise d'après les constats sur pourcentage des indices de qualité (Complétude, Cohérence et validité)

Exploration :

- ✓ Méthodes On a procédé via le script Python de savoir les différents attributs d'entrées et de sorties à partir les fichiers CSV fournit. Et les informations préliminaires qui concernent la moyenne et écart type (etc...).

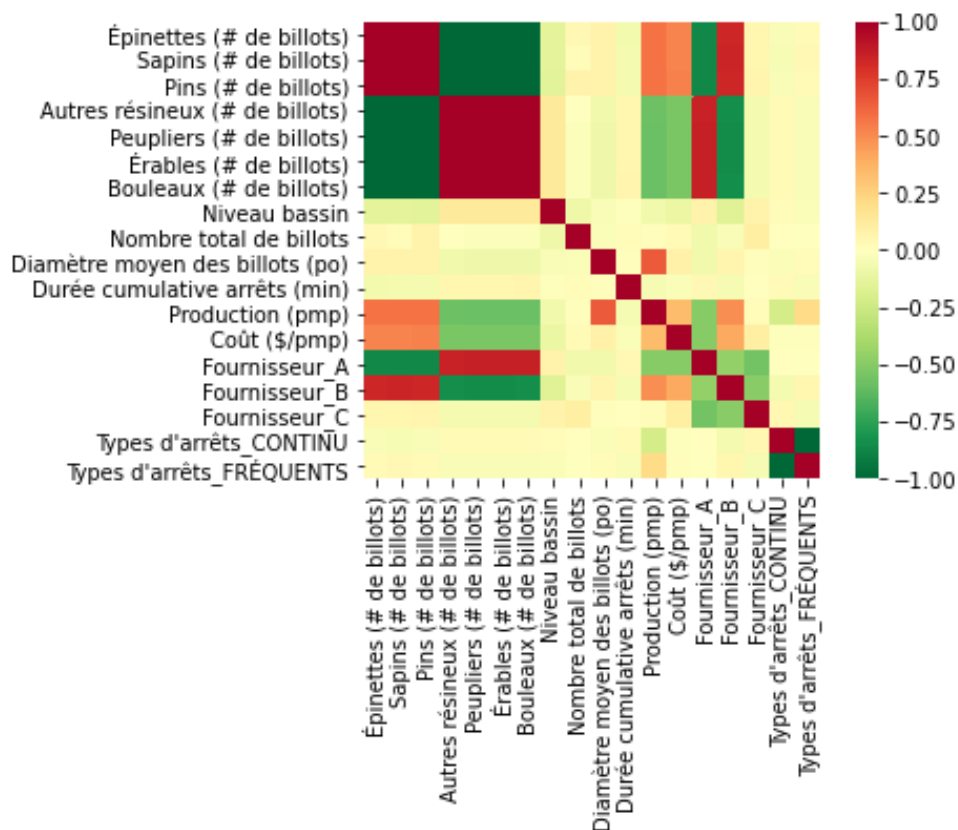
Variables et facteurs influençant la productivité

D'après le script Python, on a calculé la matrice de corrélation entre les différentes variables.

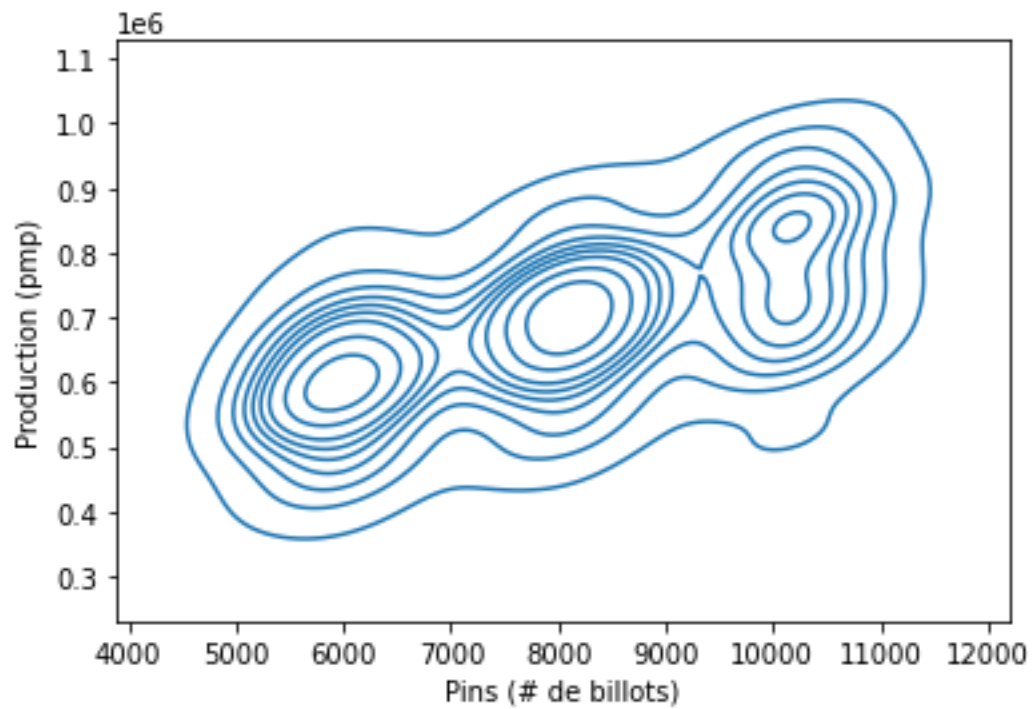
D'après cette matrice, par exemple on trouve que la productivité est corrélation fort positive avec le diamètre moyen des billots, en corrélation aussi avec variable d'entrée « Fournisseur B ». Et de même en corrélation avec le cout(\$/pmp) et la variable Épinettes, Sapins et Pins (etc...) ci-dessous un tableau qui résume les différentes corrélations entre la production et les autres variables qui concerne cette corrélation.

(2)	Fourinsseur B	Diamètre moyen des billots (po)	Bouleaux (# de billots)	Érables (# de billots)	Peupliers (# de billots)	Autres résineux (# de billots)	Pins (# de billots)	Sapins (# de billots)	Épinettes (# de billots)
Production En (pmp) (1)	0.35	0.66	- 0.58	- 0.58	- 0.58	-0.57	0.58	0.58	0.579
Facteur de corrélation entre (1) et (2)									

Ce qui conclut que la variable de sortie production est influencée par ces variables. Ci-joint la matrice de corrélation.



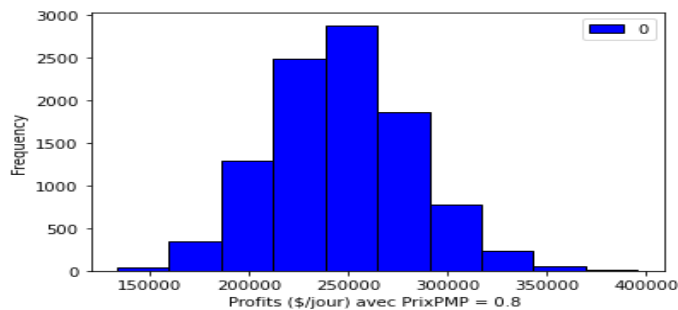
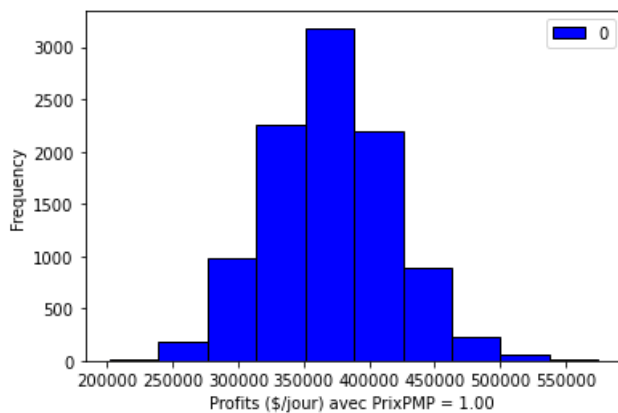
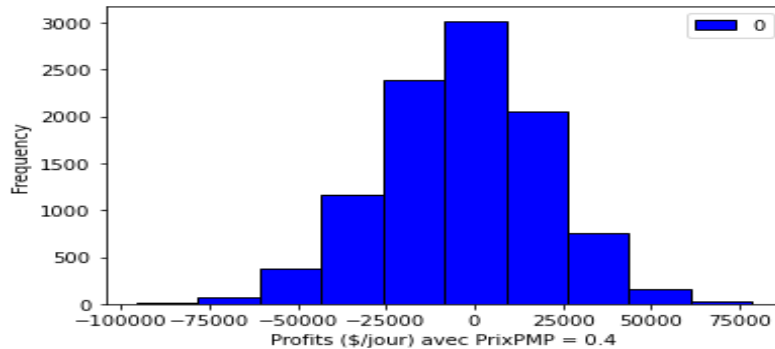
A titre exemple ici les lignes de niveaux illustre la tendance de corrélation entre 'Pins' et 'Production', sur la figure ci-dessous (corrélation entre 'Pins' et 'Production')



On peut faire la même chose avec les autres combinaisons possible entre la production et les autres variables de corrélation.

Analyse de sensibilité (Monté Carlo) (code script python)

Impact du prix du bois (\$ / pmp) sur la distribution des profits



D'après la simulation de sensibilité Monté Carlo, on observe selon les histogrammes de distribution de profit, que le profit moyen est intéressant et augmente avec l'augmentation du prix de PMP.

Note technique

D'après l'étude empirique et exploratoire des données ci-dessus. On recommande dans le but d'avoir une meilleure production à cette usine de planifier des contrats futurs avec le fournisseur B. D'autre part ciblé un type bien précise des arbres comme : Épinettes, Sapins et Pins versus d'autres ou il est conseillé de baisser le nombre (en Billots) dont le genre : Érables, Peupliers, Autres résineux. Ciblé aussi le diamètre plus grand de billots. D'autre part pour réaliser un bon profit il faut chercher des marchés avec des prix en PMP plus élevé.