

420-A55-SF ANALYSE EXPLORATOIRE DES DONNÉES groupe 12113

TP 02

Étudiant:

Mohamed Toros
Matricule : 2097154

Travail présenté au :
Professeur

Pierre-Marc Juneau

Date : 22 Avril 2021

TABLE DES MATIÈRES


Q0 - Préparation et prétraitement des données	4
Nettoyage :	4
Exploration :	Error! Bookmark not defined.
Q1 - Corrélations et régressions pour les facteurs influençant le plus la production	5
Q2-Variations de la transmittance et du contenu en sucre de la sève d'une année à l'autre	18
Q3-Modèle pour évaluer par caméra la transparence du sirop (en ligne)	22
Note technique	25

Introduction et mise en contexte

Le projet d'analyse exploratoire sujet de notre étude concerne une érablière. Ainsi, son besoin d'affaire réside dans la compréhension de l'impact des paramètres (sève, contenu de sucre du sirop, couleur sirop...etc) qui influencent la production de l'entreprise (quantité du sirop produite et la qualité associée). Donc, notre mission est de bâtir un modèle fiable de prédiction qui garantit une meilleure production avec une qualité appropriée à base des concepts statistiques, descriptives inférentielles (régression multilinéaire et méthodes multivariées).

Q0 - Préparation et prétraitement des données

Q0.1 Prétraitement : Après le chargement de notre dataset brute composé de (1875 entrées, 50 variables) On a commencé par le calcul des indices de qualité, à savoir les degrés.

<p>Donc puisque les valeurs sont presque égales à 1. On a pris la décision de supprimer les valeurs manquantes. Et aussi on a éliminé les entrées avec classes sirop égale à 0. Car d'après la phase de l'exploration il s'est avéré que toutes les entrées correspondantes sont nulles aussi donc a fait le nettoyage par la suppression de ces entrées.</p>	 <table border="1"> <thead> <tr> <th>Index</th> <th>0</th> </tr> </thead> <tbody> <tr><td>Date</td><td>1</td></tr> <tr><td>Année</td><td>1</td></tr> <tr><td>Jour Calendrier Saison</td><td>1</td></tr> <tr><td>Temp max.(°C)</td><td>0.998933</td></tr> <tr><td>Temp min.(°C)</td><td>1</td></tr> <tr><td>Temp moy.(°C)</td><td>0.998933</td></tr> <tr><td>Diff Temp (°C)</td><td>1</td></tr> <tr><td>Précip. tot. (mm)</td><td>0.9984</td></tr> <tr><td>Précip. Tot. Hiver (mm)</td><td>1</td></tr> <tr><td>Nombre épisodes gel/dégel</td><td>1</td></tr> <tr><td>Production moyenne par entaille (L)</td><td>1</td></tr> <tr><td>Débit sève (L/j)</td><td>1</td></tr> </tbody> </table>	Index	0	Date	1	Année	1	Jour Calendrier Saison	1	Temp max.(°C)	0.998933	Temp min.(°C)	1	Temp moy.(°C)	0.998933	Diff Temp (°C)	1	Précip. tot. (mm)	0.9984	Précip. Tot. Hiver (mm)	1	Nombre épisodes gel/dégel	1	Production moyenne par entaille (L)	1	Débit sève (L/j)	1
Index	0																										
Date	1																										
Année	1																										
Jour Calendrier Saison	1																										
Temp max.(°C)	0.998933																										
Temp min.(°C)	1																										
Temp moy.(°C)	0.998933																										
Diff Temp (°C)	1																										
Précip. tot. (mm)	0.9984																										
Précip. Tot. Hiver (mm)	1																										
Nombre épisodes gel/dégel	1																										
Production moyenne par entaille (L)	1																										
Débit sève (L/j)	1																										
<pre>diff= donneesCabneSucreClean1["Temp max.(°C)"] - donneesCabneSucreClean1["Temp min.(°C)"] - donneesCabneSucreClean1["Diff Temp (°C)"] diff = round(diff,2) NNC=sum(i != 0 for i in diff) DegCoherence=(NR-NNC)/NR diff=pd.DataFrame(diff) diff.index = donneesCabneSucreClean1.index</pre>	<p>On a obtenu un degré de cohérence égale à 1.</p>																										

Q1 - Corrélations et régressions pour les facteurs influençant le plus la production

Variable dépendante	Y
Q1.1	Débit de sève (L/j)
Q1.2	Sucre dans la sève (%)
Q1.3	Transmittance du sirop (grade) (%)
Q1.4	La productivité en sève par saison (L/entaille)

Afin d'établir un modèle régression multilinéaire, permettant d'expliquer les variables dépendantes ci-dessus :

Il faut tout d'abord cerner les variables explicatives en lien avec chaque variable Y.

Effectivement, une exploration préliminaire après la phase de prétraitement des données brutes (`donneesCabneSucre = pd.read_csv('CabaneASucrev0r2.csv')`) et ensuite comprendre les liens avec ces dernières, il faut filtrer nos régresseurs (par exemples ceux qu'ont en lien avec les Pixels il faut l'enlever à ce niveau et garder les autres régresseurs).

○ Q1.1 - Modèle pour : Y = Débit de sève (L/j)

En utilisant le script python me permet d'enlever les colonnes en les variables en lie avec les entités des 'Pixels' pour les couleurs RGB.

```
donneesCabneSucre2 =
donneesCabneSucre1[donneesCabneSucre1.columns.drop(list(donneesCabneSucre1.filter(regex= 'Pixel')))]
```

Après ça on suit le processus de produire le dit modèle selon les étapes dresser dans le tableau ci-dessous :

Étape 0 : <code>X = donneesCabneSucre2</code> <code>X = X.drop("Débit sève (L/j)",1)</code> <code>X = X.drop("Date",1)</code> <code>X = X.drop("Année",1)</code> <code>X = X.drop("Classe Sirop",1)</code> <code>Y = donneesCabneSucre2["Débit sève (L/j)"]</code>	Préparation de notre simulation : A ce niveau, on doit enlever les autres variables comme date et la variable catégoriale 'Classe Sirop', et préciser notre variable à expliquer sujet d'étude (Y → Débit sève (L/j))
Étape 1 (itération 1) <code>import statsmodels.api as sm</code>	Exécution le script de simulation du modèle: A ce stade on initialise la simulation de notre modèle pour la première fois (1 ^{ère} itération) avec l'ensemble

<pre> modele = sm.OLS(Y,X.assign(const = 1)) resultats = modele.fit() Y_chap = resultats.predict(X.assign(const = 1)) resultats.summary() </pre>	<p>initiale des régresseurs stocké dans la variable X, voir l'inventaire des résultats obtenus (objet fig. 1). Ensuite, on élimine les régresseurs dont les coefficients β dépassent 5% (ici : $P > t$ supérieure à 0.05 c'est l'objet des variables encadré dans la figure1). Et on garde les autres variables pour simulation futur a une 2ème itération (résultats de synthèse voir figure2). Pour les régresseurs qui restent candidats pour la 2eme itération est noté par (new_features) dans notre script Python.</p>
<p>Étape 2 (itération 2)</p>	<p>On refait la simulation pour une 2ème itération avec seulement les régresseurs restantes de l'étape 1. Ainsi, les résultats de synthèses (objet du figure 2). On examine les valeurs de β pour chaque variable, il apparait qu'il faut Éliminé à nouveau deux variables il s'agit de 'Jour calendrier saison' et 'Transmittance produit'.</p>
<p>Étape 3 (itération 3)</p>	<p>On suit le même processus comme a l'étape 2. Mais on observe que notre modèle est finalement est stabilisé car toutes les variables ont un β inférieur à 0.05. Ainsi les variables qui expliquent note variable a expliqué ou de sortie "Débit sève (L/j)" est l'ensembles des variables. {['Temp max.(°C)', 'Temp min.(°C)', 'Précip. Tot. Hiver (mm)', 'Précip. tot. (mm)', 'Production moyenne par entaille (L)', 'Sucre sève (%)', 'Temps bouilloire (h)', 'Quantité de sirop obtenue (L)']} Par contre le calcul de VIF montre autres informations aussi.</p>
<p>Étape 4</p>	<p>A ce stade pour avoir plus de visibilité sur notre modèle. Afin de réduire de plus en plus le nombre des régresseurs, et faire face aux problèmes probables en lien des fois avec la multicolniéarité entre les régresseurs. Dans cette perspective, on calcule le facteur d'inflation de variance (VIF). Effectivement, si on utilise l'ensemble des régresseurs obtenus à la 3ème itération. Pour simuler à quel ordre chaque variable x_i est corrélé par rapport aux autres. La figure 4 donne les résultats des valeurs selon l'ordre de corrélation.</p>
<p>Étape 5</p>	<p>On simule notre modèle de synthèse (avec les coffecients β) après l'élimination des variables avec VIF supérieure à 10. Mais le compromis ici on perde la qualité car le facteur R-squared est devient loin de 1 Donc on garde l'ensemble final obtenu à la simulation du modèle à l'itération 3.</p>

Régresseurs finaux est l'ensemble	{['Temp max.(°C)','Temp min.(°C)','Précip. Tot. Hiver (mm)','Précip. tot. (mm)','Production moyenne par entaille (L)','Sucre sève (%)','Temps bouilloire (h)','Quantité de sirop obtenue (L)']}
-----------------------------------	---

OLS Regression Results						
Dep. Variable:	Débit sève (L/j)	R-squared:	0.988			
Model:	OLS	Adj. R-squared:	0.988			
Method:	Least Squares	F-statistic:	7945.			
Date:	Thu, 22 Apr 2021	Prob (F-statistic):	0.00			
Time:	10:18:43	Log-Likelihood:	-12060.			
No. Observations:	1592	AIC:	2.416e+04			
Df Residuals:	1574	BIC:	2.425e+04			
Df Model:	17					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Jour Calendrier Saison	-3.8301	0.784	-4.887	0.000	-5.368	-2.293
Temp max.(°C)	-901.2569	197.063	-4.573	0.000	-1287.791	-514.723
Temp min.(°C)	-915.0067	197.143	-4.641	0.000	-1301.696	-528.317
Temp moy.(°C)	1821.2383	393.483	4.629	0.000	1049.432	2593.044
Diff Temp (°C)	13.7498	2.103	6.538	0.000	9.625	17.875
Précip. tot. (mm)	-18.2440	7.323	-2.491	0.013	-32.608	-3.880
Précip. Tot. Hiver (mm)	-0.3334	0.151	-2.212	0.027	-0.629	-0.038
Nombre épisodes gel/dégel	-16.0965	12.094	-1.331	0.183	-39.819	7.626
Production moyenne par entaille (L)	12.3819	5.032	2.461	0.014	2.512	22.252
Sucre sève (%)	-632.6478	141.500	-4.471	0.000	-910.196	-355.100
Alimentation osmoseur (L/j)	-1.5702	9.279	-0.169	0.866	-19.771	16.631
Osmoseur (heures opération/j)	1041.6631	4123.041	0.253	0.801	-7045.567	9128.893
Pression osmoseur (bar)	779.5172	1170.293	0.666	0.505	-1515.980	3075.015
Sucre sortie osmoseur (%)	-760.4954	1170.216	-0.650	0.516	-3055.841	1534.850
Température Bouilloire (°C)	-270.5755	413.686	-0.654	0.513	-1082.010	540.859
Temps bouilloire (h)	723.8918	35.591	20.339	0.000	654.082	793.702
Sucre du sirop obtenu (%)	68.2074	83.139	0.820	0.417	-94.867	231.282
Quantité de sirop obtenue (L)	13.7961	0.658	20.955	0.000	12.505	15.087
Transmittance produit (%)	-13.5656	1.246	-10.888	0.000	-16.009	-11.122
const	48.1254	73.141	0.658	0.511	-95.338	191.589
Omnibus:	187.950	Durbin-Watson:	1.617			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1607.269			
Skew:	0.177	Prob(JB):	0.00			
Kurtosis:	7.910	Cond. No.	6.28e+18			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[2] The smallest eigenvalue is 2.03e-28. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.						

Figure 1 : Résultat simulation modèle de régression multivariés (1ère itération)

OLS Regression Results

Dep. Variable:	Débit sève (L/j)	R-squared:	0.982
Model:	OLS	Adj. R-squared:	0.982
Method:	Least Squares	F-statistic:	8820.
Date:	Thu, 22 Apr 2021	Prob (F-statistic):	0.00
Time:	11:24:06	Log-Likelihood:	-12398.
No. Observations:	1592	AIC:	2.482e+04
Df Residuals:	1581	BIC:	2.488e+04
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Jour Calendrier Saison	-1.8456	0.958	-1.926	0.054	-3.725	0.034
Temp max.(°C)	39.1300	4.300	7.070	0.000	27.132	49.128
Temp min.(°C)	-22.2438	3.751	-5.930	0.000	-29.602	-14.886
Précip. Tot. Hiver (mm)	-0.5599	0.185	-3.028	0.002	-0.922	-0.197
Précip. tot. (mm)	-39.6929	7.482	-5.305	0.000	-54.369	-25.017
Production moyenne par entaille (L)	12.1921	3.801	3.208	0.001	4.738	19.647
Sucre sève (%)	-1347.5373	141.361	-9.533	0.000	-1624.812	-1070.262
Temps bouilloire (h)	397.2502	40.768	9.744	0.000	317.284	477.216
Quantité de sirop obtenue (L)	28.0430	0.518	54.098	0.000	27.026	29.060
Transmittance produit (%)	2.1842	1.373	1.591	0.112	-0.508	4.877
const	2776.7580	389.945	7.121	0.000	2011.895	3541.621

Omnibus:	462.973	Durbin-Watson:	1.528
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4894.945
Skew:	1.045	Prob(JB):	0.00
Kurtosis:	11.332	Cond. No.	1.81e+04

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.81e+04. This might indicate that there are strong multicollinearity or other numerical problems.

On enlève à nouveau les variables :

Jour calendrier saison

Et **Transmittance**

produit avec des coefficients β dépassent 5%. (0.112 > 0.5 0.054 >

Figure 2 : Résultat simulation modèle de régression multivariés (2ère itération)

<class 'statsmodels.iolib.summary.Summary'>

OLS Regression Results

Dep. Variable:	Débit sève (L/j)	R-squared:	0.982
Model:	OLS	Adj. R-squared:	0.982
Method:	Least Squares	F-statistic:	1.098e+04
Date:	Thu, 22 Apr 2021	Prob (F-statistic):	0.00
Time:	11:45:49	Log-Likelihood:	-12402.
No. Observations:	1592	AIC:	2.482e+04
Df Residuals:	1583	BIC:	2.487e+04
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Temp max.(°C)	32.4993	4.352	7.468	0.000	23.964	41.035
Temp min.(°C)	-24.1038	3.666	-6.574	0.000	-31.295	-16.912
Précip. Tot. Hiver (mm)	-0.5783	0.185	-3.124	0.002	-0.941	-0.215
Précip. tot. (mm)	-39.0255	7.490	-5.210	0.000	-53.718	-24.333
Production moyenne par entaille (L)	12.6657	3.804	3.329	0.001	5.204	20.128
Sucre sève (%)	-1343.7600	141.543	-9.494	0.000	-1621.392	-1066.128
Temps bouilloire (h)	384.3582	39.640	9.696	0.000	306.606	462.111
Quantité de sirop obtenue (L)	28.1819	0.505	55.787	0.000	27.191	29.173
const	2781.5147	374.550	7.426	0.000	2046.848	3516.181

Omnibus:	482.003	Durbin-Watson:	1.539
Prob(Omnibus):	0.000	Jarque-Bera (JB):	5265.042
Skew:	1.090	Prob(JB):	0.00
Kurtosis:	11.638	Cond. No.	1.74e+04

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.74e+04. This might indicate that there are strong multicollinearity or other numerical problems.

R-squared est pertinent

Les coefficients β sont tous significatifs

Figure 3 : Résultat simulation modèle de régression multivariés (3ère itération)

Index	VIF
Temps bouilloire (h)	57.4398
Sucre sève (%)	39.5461
Quantité de sirop obtenue (L)	28.5009
Précip. Tot. Hiver (mm)	20.1143
Production moyenne par entaille (L)	17.6212
Temp min.(°C)	7.13045
Temp max.(°C)	5.22361
Précip. tot. (mm)	2.25358

On remarque que le VIF ici dépasse 10. Donc on élimine les variables en question et on simule notre modèle à nouveau

Malgré que les tous les β sont significatifs, mais le facteur R-squared devient plus de 1 versus celui obtenu à l'itération 3

Figure 4 : Calcul de VIF (avec les variables X_i de l'itération 3)

```
<class 'statsmodels.iolib.summary.Summary'>
=====
                        OLS Regression Results
=====
Dep. Variable:          Débit sève (L/j)      R-squared:                0.355
Model:                  OLS                  Adj. R-squared:           0.354
Method:                 Least Squares        F-statistic:              291.5
Date:                   Thu, 22 Apr 2021      Prob (F-statistic):       9.65e-151
Time:                   14:42:53             Log-Likelihood:          -15264.
No. Observations:       1592                AIC:                     3.054e+04
Df Residuals:           1588                BIC:                     3.056e+04
Df Model:               3
Covariance Type:        nonrobust
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
Temp max.(°C)    587.5336    21.674     27.107    0.000    545.020    630.047
Temp min.(°C)   -340.9711    20.012    -17.039    0.000   -380.223   -301.719
Précip. tot. (mm) -42.4819    15.768     -2.694    0.007    -73.410   -11.554
const           -887.5262   230.590     -3.849    0.000   -1339.818   -435.234
=====
Omnibus:          78.889    Durbin-Watson:           1.283
Prob(Omnibus):    0.000    Jarque-Bera (JB):         96.543
Skew:             0.505    Prob(JB):                 1.09e-21
Kurtosis:         3.660    Cond. No.                 29.7
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
=====
```

Les coefficients β sont très significatifs.

Figure 5 : Modèle simulé à l'étape 4

○ Q1.2 - Modèle pour : Y = Sucre dans la sève (%)

Donc on suit les démarches comme le cas Q1.1. On trouve les résultats suivants :

OLS Regression Results						
Dep. Variable:	Sucre sève (%)	R-squared:	0.960			
Model:	OLS	Adj. R-squared:	0.960			
Method:	Least Squares	F-statistic:	2223.			
Date:	Thu, 22 Apr 2021	Prob (F-statistic):	0.00			
Time:	15:01:58	Log-Likelihood:	1693.9			
No. Observations:	1592	AIC:	-3352.			
Df Residuals:	1574	BIC:	-3255.			
Df Model:	17					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Jour Calendrier Saison	-0.0002	0.000	-1.243	0.214	-0.000	0.000
Temp max.(°C)	-0.0267	0.035	-0.760	0.448	-0.096	0.042
Temp min.(°C)	-0.0266	0.035	-0.757	0.449	-0.095	0.042
Temp moy.(°C)	0.0534	0.070	0.762	0.446	-0.084	0.191
Diff Temp (°C)	-8.276e-05	0.000	-0.219	0.826	-0.001	0.001
Précip. tot. (mm)	-0.0490	0.000	-120.833	0.000	-0.050	-0.048
Précip. Tot. Hiver (mm)	-0.0006	2.19e-05	-27.704	0.000	-0.001	-0.001
Nombre épisodes gel/dégel	0.0474	0.002	26.689	0.000	0.044	0.051
Production moyenne par entaille (L)	0.0006	0.001	0.663	0.507	-0.001	0.002
Débit sève (L/j)	-1.982e-05	4.43e-06	-4.471	0.000	-2.85e-05	-1.11e-05
Alimentation osmoseur (L/j)	0.0022	0.002	1.356	0.175	-0.001	0.005
Osmoseur (heures opération/j)	-0.9942	0.729	-1.363	0.173	-2.425	0.437
Pression osmoseur (bar)	-0.0822	0.207	-0.397	0.692	-0.489	0.324
Sucre sortie osmoseur (%)	0.0824	0.207	0.398	0.691	-0.324	0.489
Température Bouilloire (°C)	0.0527	0.073	0.719	0.472	-0.091	0.196
Temps bouilloire (h)	-0.0035	0.007	-0.495	0.621	-0.017	0.010
Sucre du sirop obtenu (%)	-0.0135	0.015	-0.918	0.359	-0.042	0.015
Quantité de sirop obtenue (L)	0.0009	0.000	6.578	0.000	0.001	0.001
Transmittance produit (%)	0.0002	0.000	0.723	0.470	-0.000	0.001
const	-0.0051	0.013	-0.397	0.691	-0.031	0.020
Omnibus:	597.643	Durbin-Watson:	2.062			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	80.164			
Skew:	0.046	Prob(JB):	3.91e-18			
Kurtosis:	1.905	Cond. No.	1.71e+19			

Notes:
 [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
 [2] The smallest eigenvalue is 1.83e-28. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

Figure 6. Résultat de simulation du modèle (1ère itération)

```

<class 'statsmodels.iolib.summary.Summary'>
=====
                        OLS Regression Results
=====
Dep. Variable:          Sucre sève (%)      R-squared:                0.960
Model:                  OLS                Adj. R-squared:          0.960
Method:                 Least Squares       F-statistic:             7555.
Date:                   Thu, 22 Apr 2021    Prob (F-statistic):      0.00
Time:                   15:10:25           Log-Likelihood:         1687.8
No. Observations:      1592              AIC:                    -3364.
Df Residuals:          1586              BIC:                    -3331.
Df Model:               5
Covariance Type:       nonrobust
=====
                        coef      std err      t      P>|t|      [0.025      0.975]
-----
Précip. tot. (mm)      -0.0489      0.000    -128.824    0.000    -0.050    -0.048
Précip. Tot. Hiver (mm) -0.0006      1.69e-05    -35.304    0.000    -0.001    -0.001
Nombre épisodes gel/dégel 0.0487      0.000     103.933    0.000      0.048      0.050
Débit sève (L/j)      -2.423e-05    3.38e-06    -7.163     0.000   -3.09e-05   -1.76e-05
Quantité de sirop obtenue (L) 0.0008      0.000      7.082     0.000      0.001      0.001
const                  1.9196      0.015     131.333    0.000      1.891      1.948
=====
Omnibus:                619.648    Durbin-Watson:           2.058
Prob(Omnibus):          0.000    Jarque-Bera (JB):        81.074
Skew:                   0.048    Prob(JB):                2.48e-18
Kurtosis:               1.899    Cond. No.:               3.73e+04
=====
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.73e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
=====

```

Tous les β sont
Significatifs
avec R-squared
pertinent aussi.

Figure 5. Résultat de simulation du modèle (2ère itération)

Mais ça n'empêche pas d'examiner la multicollinéarité comme avant. En calculant les valeurs de VIF. Le figure ci-dessous illustre les résultats suivants :

Index	VIF
Quantité de sirop obtenue (L)	72.5097
Débit sève (L/j)	71.5026
Nombre épisodes gel/dégel	12.4385
Précip. Tot. Hiver (mm)	11.5304
Précip. tot. (mm)	1.39386

Figure 7. Valeurs de VIF à l'itération 2.

Si on supprime juste les deux premières variables par exemple, on obtient plus de visibilité, après la simulation du modèle encore une fois.

```
<class 'statsmodels.iolib.summary.Summary'>
=====
                        OLS Regression Results
=====
```

Dep. Variable:	Sucre sève (%)	R-squared:	0.958
Model:	OLS	Adj. R-squared:	0.958
Method:	Least Squares	F-statistic:	1.220e+04
Date:	Thu, 22 Apr 2021	Prob (F-statistic):	0.00
Time:	15:20:32	Log-Likelihood:	1662.4
No. Observations:	1592	AIC:	-3317.
Df Residuals:	1588	BIC:	-3295.
Df Model:	3		
Covariance Type:	nonrobust		

```
=====
```

	coef	std err	t	P> t	[0.025	0.975]
Précip. tot. (mm)	-0.0497	0.000	-136.168	0.000	-0.050	-0.049
Précip. Tot. Hiver (mm)	-0.0006	1.71e-05	-35.825	0.000	-0.001	-0.001
Nombre épisodes gel/dégel	0.0500	0.000	119.699	0.000	0.049	0.051
const	1.9004	0.015	130.565	0.000	1.872	1.929

```
=====
```

Omnibus:	824.469	Durbin-Watson:	2.064
Prob(Omnibus):	0.000	Jarque-Bera (JB):	88.020
Skew:	0.062	Prob(JB):	7.70e-20
Kurtosis:	1.855	Cond. No.	4.31e+03

```
=====
```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 4.31e+03. This might indicate that there are strong multicollinearity or other numerical problems.

```
=====
```

Figure 8. Résultats de synthèse à l'itération 3 (via le calcul de VIF)

Si on calcul les valeurs de VIF de nouveau pour les régresseurs restant on obtient :

Index	VIF
Précip. Tot. Hiver (mm)	9.72708
Nombre épisodes gel/dégel	9.50231
Précip. tot. (mm)	1.2509

Figure 9. Calcul des valeurs de VIF après l'itération 4.

Donc, on remarque ici, qu'un problème de multicollinéarité est réglé à ce nouveau par l'élimination des dites variables. Ce qui montre l'intérêt de calcul de VIF comme boussole pour avancer la recherche d'un modèle optimal.

Régresseurs finaux est l'ensemble	{'Précip. Tot. Hiver (mm)', 'Précip. tot. (mm)', 'Nombre épisodes gel/dégel'}
-----------------------------------	--

○ Q1.3 - Modèle pour : Y = % de transmittance du sirop (grade)

OLS Regression Results						
=====						
Dep. Variable:	Transmittance produit (%)	R-squared:	0.382			
Model:	OLS	Adj. R-squared:	0.375			
Method:	Least Squares	F-statistic:	57.19			
Date:	Thu, 22 Apr 2021	Prob (F-statistic):	1.16e-150			
Time:	15:35:48	Log-Likelihood:	-5792.3			
No. Observations:	1592	AIC:	1.162e+04			
Df Residuals:	1574	BIC:	1.172e+04			
Df Model:	17					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Jour Calendrier Saison	-0.1622	0.015	-10.922	0.000	-0.191	-0.133
Temp max.(°C)	-9.2035	3.863	-2.382	0.017	-16.781	-1.626
Temp min.(°C)	-9.1784	3.865	-2.375	0.018	-16.760	-1.597
Temp moy.(°C)	18.3111	7.715	2.374	0.018	3.179	33.443
Diff Temp (°C)	-0.0251	0.042	-0.603	0.547	-0.107	0.056
Précip. tot. (mm)	0.0771	0.143	0.539	0.590	-0.204	0.358
Précip. Tot. Hiver (mm)	-0.0016	0.003	-0.528	0.597	-0.007	0.004
Nombre épisodes gel/dégel	-0.1371	0.236	-0.581	0.561	-0.600	0.326
Production moyenne par entaille (L)	0.0722	0.098	0.735	0.463	-0.121	0.265
Débit sève (L/j)	-0.0052	0.000	-10.888	0.000	-0.006	-0.004
Sucre sève (%)	2.0076	2.778	0.723	0.470	-3.440	7.456
Alimentation osmoseur (L/j)	-0.1966	0.181	-1.087	0.277	-0.552	0.158
Osmoseur (heures opération/j)	93.1801	80.404	1.159	0.247	-64.530	250.890
Pression osmoseur (bar)	-36.1778	22.816	-1.586	0.113	-80.931	8.576
Sucre sortie osmoseur (%)	36.2767	22.815	1.590	0.112	-8.473	81.027
Température Bouilloire (°C)	13.3830	8.065	1.659	0.097	-2.436	29.202
Temps bouilloire (h)	2.0558	0.779	2.641	0.008	0.529	3.583
Sucre du sirop obtenu (%)	-2.7918	1.621	-1.722	0.085	-5.971	0.387
Quantité de sirop obtenue (L)	-0.0288	0.015	-1.988	0.047	-0.057	-0.000
const	-2.2642	1.426	-1.588	0.113	-5.061	0.533
=====						
Omnibus:	25.733	Durbin-Watson:	2.094			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	45.068			
Skew:	-0.080	Prob(JB):	1.64e-10			
Kurtosis:	3.809	Cond. No.	1.61e+19			
=====						
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[2] The smallest eigenvalue is 2.05e-28. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.						

Figure 10. Tableau de synthèse à l'itération 1

```

<class 'statsmodels.iolib.summary.Summary'>
=====
                        OLS Regression Results
=====
Dep. Variable:      Transmittance produit (%)    R-squared:      0.063
Model:              OLS                        Adj. R-squared:  0.061
Method:             Least Squares              F-statistic:    35.51
Date:               Thu, 22 Apr 2021            Prob (F-statistic): 3.29e-22
Time:               15:41:18                    Log-Likelihood: -6123.5
No. Observations:   1592                       AIC:            1.226e+04
Df Residuals:       1588                       BIC:            1.228e+04
Df Model:           3
Covariance Type:    nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
Temp max.(°C)      -0.3496      0.082      -4.263      0.000      -0.510      -0.189
Temp min.(°C)      -0.0646      0.067      -0.962      0.336      -0.196      0.067
Quantité de sirop obtenue (L)  0.0011      0.003      0.417      0.677      -0.004      0.006
const              43.9816      0.673      65.356      0.000      42.662      45.302
=====
Omnibus:            1.210      Durbin-Watson:      2.044
Prob(Omnibus):      0.546      Jarque-Bera (JB):    1.106
Skew:               0.049      Prob(JB):            0.575
Kurtosis:           3.084      Cond. No.             381.
=====
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
=====

```

Figure 11. Tableau de synthèse des valeurs à l'itération 2

On remarque qu'on a juste une seule variable qui reste avec une chute de la valeur de R-squared et loin de 1. Donc on conserve les résultats avec le modèle précédent. Ce qui valide aussi à ce niveau par le calcul de VIF, en effet on obtient les résultats suivants de VIF:

Index	VIF
Temp max.(°C)	2.20987
Quantité de sirop obtenue (L)	2.10621
Temp min.(°C)	1.66835

Figure 10. Les résultats de calcul de VIF. (Itération 01)

○ Q1.4 - Modèle pour : Y = La productivité en sève par saison (L/entaille)

Si on suit toujours les mêmes démarches de notre processus, on trouve après simulation de notre modèle à la première itération les résultats de synthèse suivants :

OLS Regression Results						
Dep. Variable:	Production moyenne par entaille (L)		R-squared:	0.945		
Model:	OLS		Adj. R-squared:	0.945		
Method:	Least Squares		F-statistic:	1594.		
Date:	Thu, 22 Apr 2021		Prob (F-statistic):	0.00		
Time:	16:00:11		Log-Likelihood:	-3624.8		
No. Observations:	1592		AIC:	7286.		
Df Residuals:	1574		BIC:	7382.		
Df Model:	17					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Jour Calendrier Saison	-0.0031	0.004	-0.775	0.439	-0.011	0.005
Temp max.(°C)	-0.6856	0.992	-0.691	0.489	-2.631	1.259
Temp min.(°C)	-0.6783	0.992	-0.684	0.494	-2.624	1.268
Temp moy.(°C)	1.3771	1.980	0.695	0.487	-2.507	5.261
Diff Temp (°C)	-0.0073	0.011	-0.684	0.494	-0.028	0.014
Précip. tot. (mm)	0.0086	0.037	0.235	0.814	-0.063	0.081
Précip. Tot. Hiver (mm)	0.0151	0.001	23.161	0.000	0.014	0.016
Nombre épisodes gel/dégel	1.8936	0.037	50.936	0.000	1.821	1.966
Débit sève (L/j)	0.0003	0.000	2.461	0.014	6.28e-05	0.001
Sucre sève (%)	0.4722	0.712	0.663	0.507	-0.924	1.868
Alimentation osmoseur (L/j)	0.0178	0.046	0.385	0.701	-0.073	0.109
Osmoseur (heures opération/j)	-7.9298	20.613	-0.385	0.701	-48.361	32.501
Pression osmoseur (bar)	5.6881	5.850	0.972	0.331	-5.786	17.163
Sucre sortie osmoseur (%)	-5.6672	5.850	-0.969	0.333	-17.141	5.807
Température Bouilloire (°C)	-2.3009	2.068	-1.113	0.266	-6.357	1.755
Temps bouilloire (h)	0.2803	0.200	1.402	0.161	-0.112	0.672
Sucre du sirop obtenu (%)	0.5004	0.416	1.204	0.229	-0.315	1.316
Quantité de sirop obtenue (L)	-0.0127	0.004	-3.436	0.001	-0.020	-0.005
Transmittance produit (%)	0.0047	0.006	0.735	0.463	-0.008	0.017
const	0.3549	0.366	0.971	0.332	-0.362	1.072
Omnibus:	56.875	Durbin-Watson:	0.063			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	29.744			
Skew:	-0.144	Prob(JB):	3.48e-07			
Kurtosis:	2.395	Cond. No.	1.76e+19			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[2] The smallest eigenvalue is 1.73e-28. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.						

Figure 11. Tableau de synthèse après simulation 1ère itération

A ce niveau on va enlever les variables avec des coefficients beta plus élevés à 5%. Et on simule encore une fois le modèle et on obtient les résultats illustrés dans le tableau de synthèse ci-dessous :


```

<class 'statsmodels.iolib.summary.Summary'>
=====
                        OLS Regression Results
=====
Dep. Variable:      Production moyenne par entaille (L)      R-squared:      0.945
Model:              OLS                                     Adj. R-squared: 0.945
Method:             Least Squares                          F-statistic:    6789.
Date:               Thu, 22 Apr 2021                       Prob (F-statistic): 0.00
Time:               16:09:06                               Log-Likelihood: -3629.4
No. Observations:   1592                                   AIC:            7269.
Df Residuals:       1587                                   BIC:            7296.
Df Model:           4
Covariance Type:    nonrobust
=====
                    coef    std err          t      P>|t|      [0.025      0.975]
-----
Précip. Tot. Hiver (mm)      0.0146      0.000     30.675      0.000      0.014      0.016
Débit sève (L/j)             0.0003     9.18e-05     3.499      0.000      0.000      0.001
Nombre épisodes gel/dégel    1.9158      0.013    146.802      0.000      1.890      1.941
Quantité de sirop obtenue (L) -0.0098      0.003     -3.177      0.002     -0.016     -0.004
const                      -22.0693      0.412    -53.615      0.000    -22.877    -21.262
=====
Omnibus:            68.279   Durbin-Watson:      0.055
Prob(Omnibus):      0.000   Jarque-Bera (JB):    32.541
Skew:               -0.133   Prob(JB):            8.59e-08
Kurtosis:           2.352   Cond. No.:           3.72e+04
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.72e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
=====

```

Figure 12. Tableau de synthèse après simulation 2ère itération

En. Basant les résultats du tableau ci-dessus. On remarque que tout est beau mais pour ce qui concerne des problèmes probables en lien avec la multicollinéarité entre les variables x_i de la matrice X , il faut examiner à ce stade la méthode du calcul de VIF. En effet, ce calcul donne comme résultat :

Index	VIF
Débit sève (L/j)	66.4155
Quantité de sirop obtenue (L)	66.4012
Nombre épisodes gel/dégel	12.113
Précip. Tot. Hiver (mm)	11.528

Figure 13. Calcul de VIF des variables qui correspond à l'itération 2

Ma stratégie versus cette situation repose à enlever d'abord les deux premières variables qui présentent une forte corrélation avec les autres variables.


```

<class 'statsmodels.iolib.summary.Summary'>
=====
                        OLS Regression Results
=====
Dep. Variable:      Production moyenne par entaille (L)      R-squared:      0.944
Model:              OLS                                     Adj. R-squared: 0.944
Method:             Least Squares                          F-statistic:    1.345e+04
Date:               Thu, 22 Apr 2021                        Prob (F-statistic): 0.00
Time:               16:24:45                                Log-Likelihood: -3637.2
No. Observations:   1592                                    AIC:            7280.
Df Residuals:       1589                                    BIC:            7297.
Df Model:           2
Covariance Type:    nonrobust
=====
                    coef    std err          t      P>|t|      [0.025    0.975]
-----
Précip. Tot. Hiver (mm)    0.0149    0.000    31.271    0.000    0.014    0.016
Nombre épisodes gel/dégel  1.9057    0.012   163.500    0.000    1.883    1.929
const                    -21.8669    0.406   -53.850    0.000   -22.663   -21.070
=====
Omnibus:              76.020    Durbin-Watson:      0.039
Prob(Omnibus):        0.000    Jarque-Bera (JB):    35.256
Skew:                 -0.145    Prob(JB):            2.21e-08
Kurtosis:             2.331    Cond. No.            4.31e+03
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.31e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

```

Figure 14. Tableau de synthèse après simulation 3^{ème} itération

Effectivement, notre modèle est optimisé avec un ensemble des régresseurs cerné a deux variables {'Précip. Tot. Hiver (mm)', 'Nombre épisode gel/dégel'}. D'autre par le calcul à nouveau les valeurs de VIF on valide une fois pour toute notre décision.

Index	VIF
Précip. Tot. Hiver (mm)	9.50156
Nombre épisodes gel/dégel	9.50156

Figure 15. Calcul de VIF des variables qui correspond à l'itération 3.

Q2-Variations de la transmittance et du contenu en sucre de la sève d'une année à l'autre

Tout d'abord pour analyse empirique on génère un graphique de boîte à moustache (boxplot) pour les années 2014, 2015 et 2016. Et pour une meilleure visibilité on procède d'abord à la suppression des valeurs aberrantes afin de ne pas biaiser notre modèle et par la suite l'analyse associée.

○ Q2.1.1 – Sucre sève

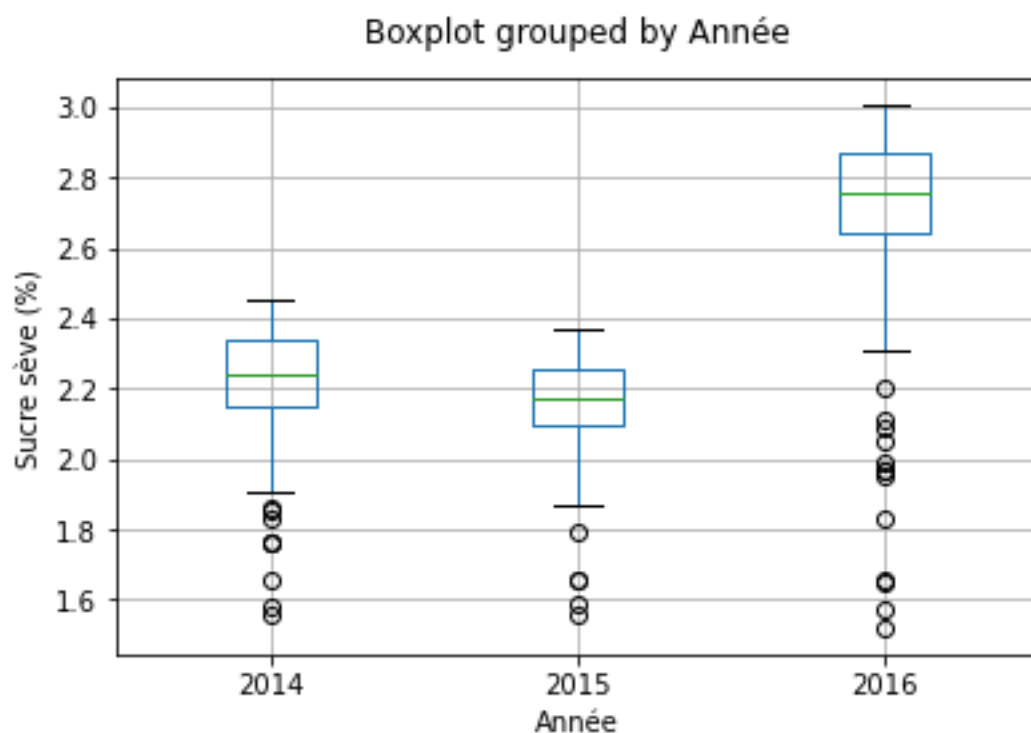


Figure 15. Boîte à moustache du pourcentage du sucre sève par année (2014,2015 et 2016).

Année	Calcul du moyen	$\mu_{\text{Année}}$
2014	sucres2014.mean()	$\mu_{2014} = 2.1928235294117644$
2015	sucres2015.mean()	$\mu_{2015} = 2.1287999999999996$
2016	sucres2016.mean()	$\mu_{2016} = 2.6519101123595505$
<p>Le calcul de P-Value pour voir la tendance globale entre les 3 années = 5.28×10^{-28}</p> <p>Comme illustré dans les résultats de la figure 16. Donc on constate que le p-value est très petite donc il y a une tendance entre les 3 années 2014, 2015 et 2016.</p>		

Inde ▲	Type	Size	Value
0	Array of float64	(1,)	[84.57979123]
1	Array of float64	(1,)	[5.28535964e-28]

Figure 16. Résultats des indices du tendance globale qui concerne le pourcentage du sucre sève entre les années (2014,2015 et 2016).

Donc ici on accepte l'hypothèse $H_0 : \mu_{2014} = \mu_{2015} = \mu_{2016}$

○ Q2.1.2 –Transmittance produit (%)

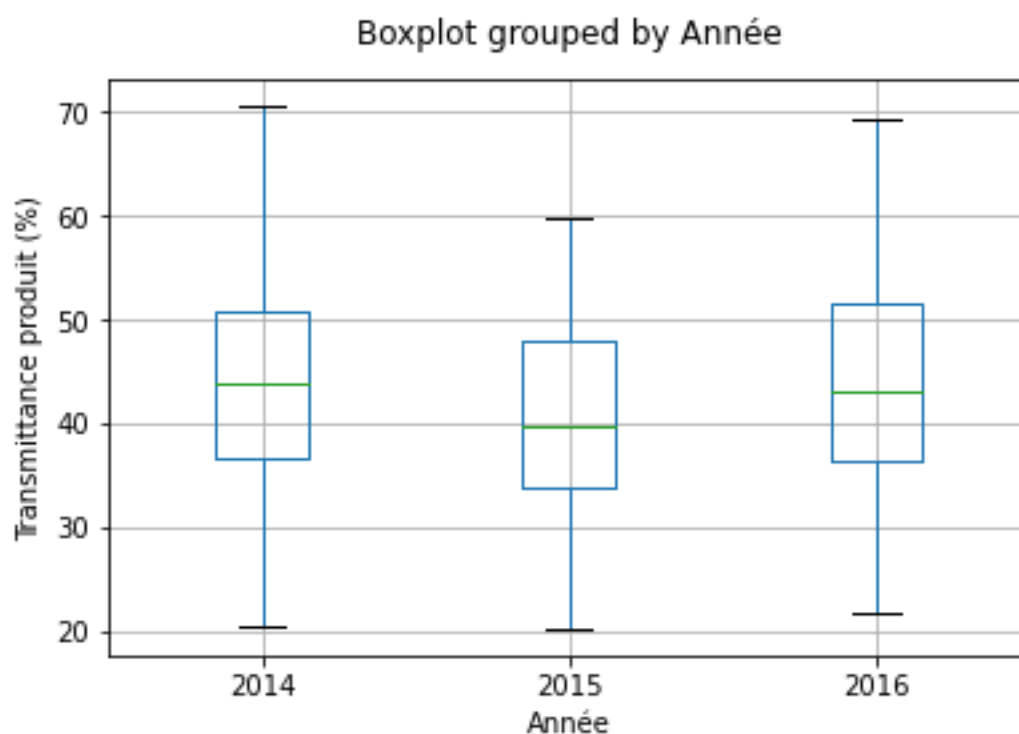


Figure 17. Boite à moustache du pourcentage du transmittance produit (%) selon les années (2014,2015 et 2016).

Année	Calcul du moyen	$\mu_{\text{Année}}$
2014	Transmittance2014.mean()	$\mu_{2014} = 44.03659090909091$
2015	Transmittance2015.mean()	$\mu_{2015} = 40.29711538461538$
2016	Transmittance2016.mean()	$\mu_{2016} = 43.914719101123595$
<p>Le calcul de P-Value pour voir la tendance globale entre les 3 années = 0.10211388 Comme illustré dans les résultats du de la figure 18. Donc on constate que le p-value n'est pas significatif car dépasse 0.05, donc il n'y a une tendance a observé entre les 3 années 2014, 2015 et 2016. Donc ici on rejette le $H_0 : \mu_{2014} = \mu_{2015} = \mu_{2016}$</p>		

`pvalue01=sts.f_oneway(Transmittance2014,Transmittance2015,Transmittance2016)`

pvalue01 - Tuple (2 elements)			
	Type	Size	Value
	Array of float64	(1,)	[2.3048579]
	Array of float64	(1,)	[0.10211388]

Figure 18. Résultats des indices du tendance globale qui concerne le pourcentage du transmittance du produit entre les années (2014,2015 et 2016).

○ **Q2.2 –**

On remarque que la différence est significative pour la transmittance, donc on suit la même façon calcul comme avant, mais cette fois-ci entre deux années (deux à deux) avec celle qui représente le plus d'écart (Année 2016).

P-Value entre deux années :	Calcul correspondant aux deux années
<p>Pvalue3=sts.ttest_ind(Transmittance2015,Transmittance2016)</p>	<div data-bbox="915 407 1256 743"> <p>pvalue03 - Tuple (2 ele</p> <p>[−10.36598683]</p> <p>[7.24404099e−20]</p> </div> <div data-bbox="1256 407 1533 701"> <p>P-Value ici est significatif donc il y a une tendance entre les deux années 2015 et 2016 pour qui concerne sucre sève. Donc on accepte L'hypothèse : $H_0 : \mu_{2014} = \mu_{2016}$</p> </div>
<p>pvalue4=sts.ttest_ind(Transmittance2014,Transmittance2016)</p>	<div data-bbox="915 806 1256 1163"> <p>pvalue04 - Tuple (2 ele</p> <p>[−9.63873199]</p> <p>[4.21037876e−17]</p> </div> <div data-bbox="1256 806 1533 1121"> <p>P-Value ici est significatif donc il y a une tendance entre les deux années 2014 et 2016 pour qui concerne sucre sève. Donc on accepte L'hypothèse $H_0 : \mu_{2015} = \mu_{2016}$</p> </div>

Q3-Modèle pour évaluer par caméra la transparence du sirop (en ligne)

L'objectif ici, est de développer un modèle de régression robuste (Fiable) d'estimer la transparence du sirop (**Y**) et de son grade (**YClass**) à partir des lecteurs des pixels des images.

Donc, on deux choses à retenir à partir les exigences des besoins d'affaires. Premièrement, garantir un modèle le plus optimisé possible avec attribut de qualité la **robustesse**.

La stratégie à suivre il faut d'abord isolé les variables x_i (Pixel*i*_ZZ avec **i = 1,..9** et **ZZ=R,G ou B**) pour obtenir une nouvelle matrice X à partir le dataset prétraité au début (le dataset prétraité et nettoyé ici dans le script on désigne par `donneesCabneSucre1`)

```
X = donneesCabneSucre1[donneesCabneSucre1.columns.drop(list(donneesCabneSucre1.filter(regex= '^((?!Pixel).)*$')))]
```

Ensuite, il faut examiner par exemple l'ordre de corrélation entre ces variables par exemple on calcul les valeurs de VIF des variables Pixel*i*_ZZ

Index	VIF	
Pixel6_R	90.8236	<p>Donc, on constate ici que les pixels (Pixeli_R , $i = 1 \dots 9$) qu'ont plus de corrélation versus les pixels des autres canaux de couleurs (Pixeli_G, Pixeli_B et , $i = 1 \dots 9$). Si on débarrasse par exemple des variables en lien avec les valeurs de canal rouge, selon moi ça va baisser l'information prise au niveau de chaque pixel, car chaque pixel est considéré comme un point atomique (besoin les trois couleurs à la fois pour s'identifier). Donc cette stratégie a écarté. En Plus on a ici une prédiction de la transparence du sirop el d'ailleurs, la prédiction de la catégorie associée comme 2ème sortie à prédire aussi à partir le modèle à concevoir. Selon les résultats théoriques vus dans le cours on cherche une combinaison linéaire qui permettrait de mieux prédire nos y_i (miniser les erreurs de prédiction) avec l'approche de PCA peut nous offrir cette opportunité d'affaire et avec PCR aussi. Effectivement, les vecteurs T_i sont orthogonaux entre eux (deux é deux) donc automatiquement, on débarrasse du problème relié au corrélation ces derniers vont être les répresseurs qui sont absolument vont réaliser un modèle robuste et fiable qui répond aux exigences citées au début. Au niveau de script accompagné, on a bâti (objet section «==Q3==») un modèle pour générer les nouvelles régresseurs à partir de les composantes de X des différents pixels.</p>
Pixel7_R	90.4361	
Pixel9_R	90.2966	
Pixel3_R	90.235	
Pixel4_R	88.5628	
Pixel2_R	88.4738	
Pixel5_R	87.8774	
Pixel1_R	86.8966	
Pixel8_R	84.2508	
Pixel6_G	17.862	
Pixel2_G	16.9984	
Pixel5_G	16.8002	
Pixel4_G	16.5401	
Pixel1_G	16.4446	
Pixel3_G	16.3406	
Pixel7_G	16.2317	
Pixel8_G	15.7413	
Pixel9_G	14.8064	
Pixel6_B	4.70376	
Pixel4_B	4.68996	
Pixel8_B	4.56021	
Pixel5_B	4.55568	
Pixel2_B	4.44906	
Pixel3_B	4.41509	
Pixel9_B	4.32483	
Pixel1_B	4.30774	
Pixel7_B	4.2769	

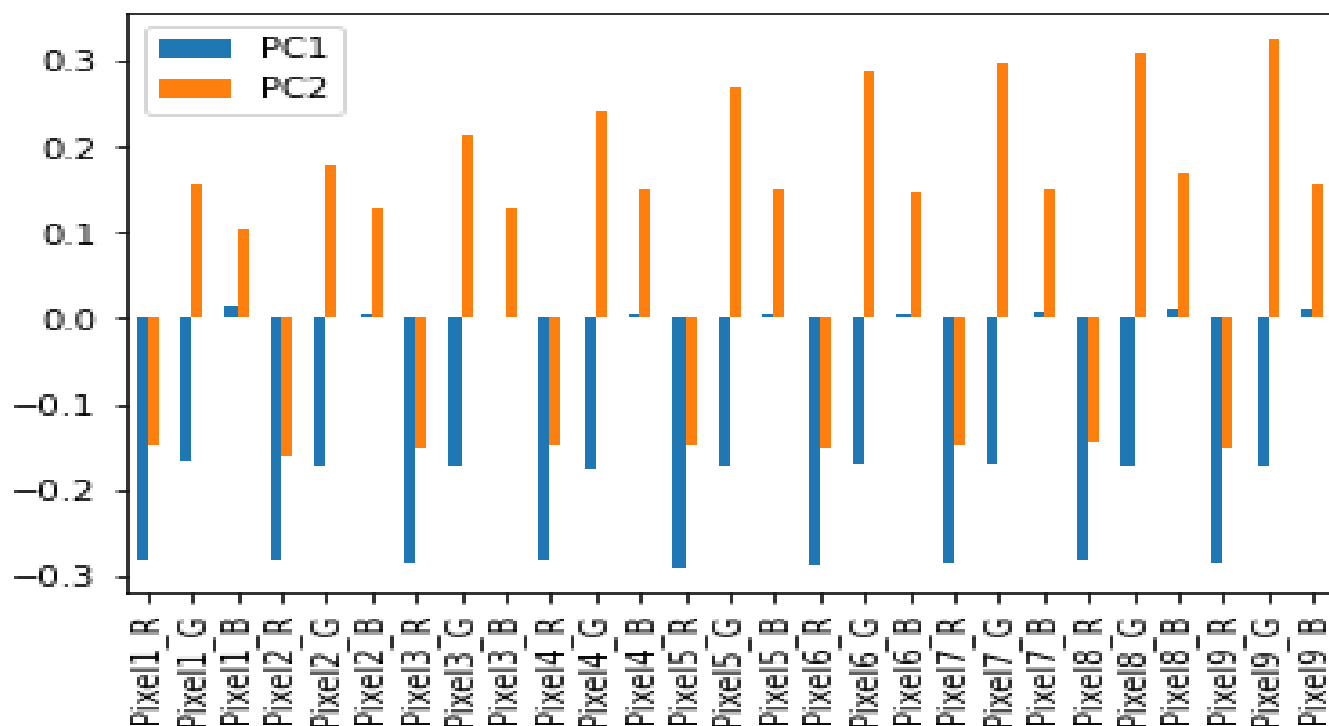


Figure 19. Analyse des composantes principales qui illustre les tendances des variables combinées

D'après le graphique ci-dessus, on observe que la 1^{ère} composante combine les différents pixels corrélés en lien avec les canaux de couleurs R et G. Par contre l'autre composante PC2 offre d'autres informations et tendances à explorer plus loin dans notre analyse.

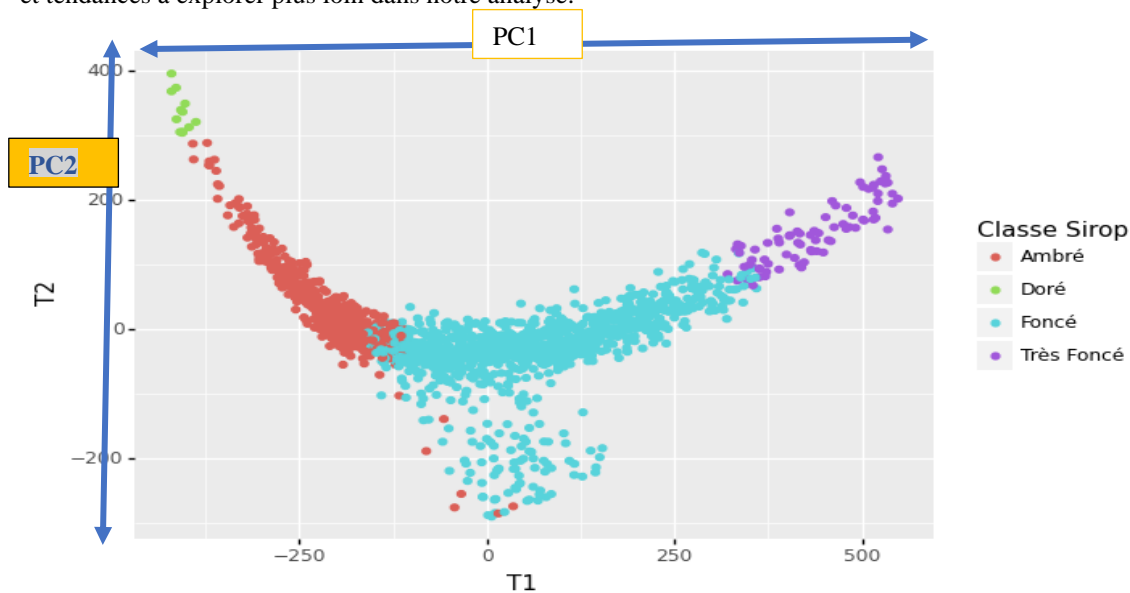


Figure 20. Graphique qui explique tendance des T_i à partir PCI

On remarque que si PC1 (Pixel_R et Pixel_G) augmente le T1 augmente aussi et les classes sirop Ambré et Doré ont même comportement et chute de T2, et une tendance contraire pour ce qui concerne les autres classes de sirop Foncé et très foncé on remarque que si PC1 évolue la même tendance pour T1 et si PC2 augmente alors T2 augmente aussi. Sachant que le sirop de la classe Foncé a un comportement un peu perturbé aux valeurs du milieu de PC1, PC2 ou T1, T2.

```
<class 'statsmodels.iolib.summary.Summary'>
=====
                        OLS Regression Results
=====
Dep. Variable:      Transmittance produit (%)      R-squared:      0.962
Model:              OLS                          Adj. R-squared: 0.962
Method:             Least Squares                 F-statistic:    2.032e+04
Date:               Thu, 22 Apr 2021               Prob (F-statistic): 0.00
Time:               21:43:09                       Log-Likelihood: -3572.0
No. Observations:   1595                          AIC:            7150.
Df Residuals:       1592                          BIC:            7166.
Df Model:            2
Covariance Type:    nonrobust
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
T1            -0.0597      0.000     -201.118      0.000     -0.060     -0.059
T2             0.0094      0.001       14.028      0.000       0.008       0.011
const         43.5841      0.057       765.504      0.000       43.472       43.696
=====
Omnibus:                 340.791    Durbin-Watson:           2.029
Prob(Omnibus):            0.000    Jarque-Bera (JB):        779.845
Skew:                     1.180    Prob(JB):                4.56e-170
Kurtosis:                  5.482    Cond. No.                 192.
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
=====
In [337]:
```

Figure 21. Tableau de synthèse de 'PCR'

On constate effectivement, que notre modèle est optimal (robuste) les coefficients β sont nuls et le R-Squared est plus pertinent car plus proche à 1.

Note technique d'analyse (conclusion et recommandations)

L'analyse effectuée au cours de ce travail, qui vise à améliorer la production du sirop et sa qualité. L'étude basée sur l'analyse des données brute depuis les années 2000. L'analyse à dévoiler une certaine corrélation entre les variables. Ainsi, on aboutit à réduire et cerner le nombre de ces variables à partir les technique de régression, et bâtir finalement les modèles de prédiction pour le dédit et le sucre de sève et la couleur du sirop en utilisant les mécanismes de régression multivariées et multilinéaire. Et basé aussi sur les résultats expérimentale (lecture valeurs des pixels RGB) via les concepts de PCA et PCR pour expliquer le comportement et les tendances observées en exploitant les graphiques générés. Donc, on recommande fortement le propriétaire de l'érablière d'adopter les modèles de prédiction et les tendances dictées pour avoir les meilleurs résultats dans le futur et réaliser des chiffres d'affaires et une marque concurrentielle dans le marché de la production du sirop avec qualité optimale.