Influența consumului de alcool al tinerilor asupra performanțelor școlare

în legătură directă cu mediul de afaceri local

**Introducere**

Unul dintre cele mai importante și vehiculate subiecte în rândul discuțiilor cu tinerii și vieții lor sociale este consumul de băuturi alcoolice. Fie că vine vorba de o petrecere, o simplă ieșire vineri seara sau un banchet, alcoolul nu lipsește de la masă. Putem pune această paradigmă pe umerii globalizării și politicii democratice populare, mai ales în rândul statelor vestice, aspecte care au contribuit enorm la crearea relațiilor economice de tip simbiotic, implicit la dezvoltarea politicilor avantajoase de import/export. Astfel, accesul la orice categorie de băuturi alcoolice este facil în zilele noastre.

În această ordine de idei, am ales să aducem în lumină influența alcoolului asupra performanțelor academice ale tinerilor, în relație strânsă cu mediul de afaceri (local). Putem afirma cu tărie faptul că orice manager sau proprietar de magazin, pub sau club dorește să afle tendința cererii în această industrie. Dacă alcoolul este și va fi parte din suita opțiunilor din cadrul oricărui eveniment social, ar putea atrage sume considerabile de capital, în lumina unor câștiguri fructoase, în schimbul unor amintiri frumoase.

Pentru acest studiu propunem spre analiză următoarele întrebări de cercetare:

1. Am putea prezice succesul producătorilor și distribuitorilor de băuturi alcoolice pe termen lung (minim 10 ani)?
2. Influențează notele consumul alcoolului?
3. Influențează consumul alcoolului performanțele academice? Dacă da, cum?

Motivația alegerii tezelor anterior menționate este următoarea:

Este important pentru proprietarii/managerii magazinelor din proximitatea școlilor să analizeze cele mai vândute produse din gama pe care o oferă pentru a solicita, la nevoie, o majorare a stocului (desigur, acest lucru se va îndeplini doar în cazul în care predicțiile sunt viabile și veritabile).

În al doilea rând, ne putem pune în pielea unui posibil investitor privat sau public căruia îi sare în atenție respectiva unitate educațională. Performanțele academice ridicate aduc de la sine o notorietate crescândă a școlii, astfel investițiile devin mai profitabile, iar investitorii, mai cunoscuți cel puțin în plan local.

Rămânând tot în zona financiară a discuției, vom privi ansamblul prin prisma resursei umane, a elevilor. Din experiența convențională colectivă putem extrage ideea că liceele și școlile cu performanțele academice cele mai ridicate au la bază, printre altele, elevi ce provin din familii cu posibilități financiare sporite. Tinerii au acces la meditații, activități extracuriculare de dezvoltare profesională și personală și, în unele cazuri, părinții lor fac parte din clasa intelectuală. Datorită acestor oportunități, un antreprenor inspirat se va orienta înspre aceste zone „calde”, pentru a intra pe piață, a-și forma o clientelă fidelă și, în cele din urmă, a crea o afacere profitabilă și prosperă.

În ultimul rând, dorim să mai subliniem un aspect vital pentru viitorul comerțului cu băuturi alcoolice. Statistic vorbind, persoanele care încep să bea alcool de la vârsta de 15-16 ani au șanse de 4 ori mai mari să consume alcool constant sau, în cazuri mai grave, să ajungă chiar în prag de alcoolism la maturitate, față de cei care încep după vârsta de 20 de ani. Deși acest lucru nu poate decât să ne întristeze, este un fapt incontestabil și real, care este decisiv în soarta companiilor din industrie. De aceea, în analiza noastră, vom lua în calcul și vârsta elevilor ce consumă cu regularitate alcool, pentru a stabili dacă tendința consumului este sau nu în creștere, pentru generațiile viitoare (tendința pe termen lung).

Cât despre alte studii de cercetare asemănătoare, am găsit 2 articole care au lucrat pe același set de date:

<http://rstudio-pubs-static.s3.amazonaws.com/346112_e11fa03b9fb34cf48b4d16bfa9b5b160.html>

<https://rstudio-pubs-static.s3.amazonaws.com/365904_90b5ff49711448b5aa6ba2f09ffb4fc3.html?fbclid=IwAR2T16fXAnjvAkiCYqPyERACy2fZDZIsKfggMO0okoPzpesnA1aba8iar_8>

Răsfoind aceste pagini, descoperim o serie de grafice, tabele și view-uri care arată diverse distribuții numerice ale variabilelor independente (spre exemplu, un pie-chart cu educația familiei sau raportul dintre sexe, nivelul de alcool consumat în timpul săptămânii în raport cu numărul de elevi etc.).

Cu toate acestea, dacă ne privim în ansamblu, observăm că proiectele se învârt în jurul următoarei paradigme: influențează sau nu alcoolul consumat performanțele academice ale elevilor? După o serie de calcule și indici, au ajuns la concluzia că da, alcoolul joacă un rol pivotal în dezvoltarea profesională armonioasă a elevilor. De asemenea, autorul a mai constatat că un consum ridicat de alcool în timpul săptămânii determină un consum ridicat în timpul weekendului, în urma folosirii funcției cor (correlation = corelație) dintre dalc (alcool consumat în timpul săptămânii) și walc (alcool consumat în weekend), unde indicele calculat a fost 0.616. Ca punct de referință, cu cât indicele este mai apropiat de 1, cu atât legătura este mai puternică.

**Setul de date**

Pentru cercetarea noastră am ales următorul set de date. <https://www.kaggle.com/datasets/uciml/student-alcohol-consumption>

Avem peste 30 de indicatori din care putem alege și peste 300 de elevi care au completat chestionarul. Atenție, elevii sunt din Portugalia, au vârsta cuprinsă între 15 și 22 de ani și studiază la una din două școli (Gabriel Pereira, respectiv Mousinho da Silveira, aflate la distanță de circa 1h și 10 minute cu mașina una de cealaltă). Fișierul student-mat.csv conține notele a 395 de elevi la matematică, pe ambele semester plus nota finală pe an, iar student-por.csv este la fel, cu mențiunea că notele sunt la limba portugheză și 649 de elevi au completat formularul.

Cu toate acestea, am descoperit un lucru surprinzător când am făcut comparație cu alte proiecte desfășurate pe același set de date sau chiar cu o concluzie pe care a împărtășit-o autorul. Ei afirmă că ar fi 382 de elevi care au completat ambele chestionare, iar noi am găsit doar 370. Explicația, în viziunea noastră, este că noi am pus toate filtrele în clauza de join, pe care l-am rulat cu ajutorul programului Tableau Prep, iar ei au scăpat câte ceva.

Observație: Primul link, la rularea printr-un program anti-plagiat, a obținut un scor de 100% la originalitate, în timp ce al doilea link doar 36%. Credem că cel din urmă a folosit ca punct de plecare primul proiect.

Din această cauză, vom continua analiza cu cei 370 de elevi pe care i-am obținut noi.

Mai jos avem avem toate coloanele și explicația aferentă:

**school** – școala unde învață elevul (‘MS’=Mousinho da Silveira sau ‘GP’=Gabriel Pereira)

**sex** – masculin sau feminin

**age** – vârsta elevului, între 15 și 22 de ani

**address** – adresa de reședința a studentului (‘R’=rural sau ‘U’=urban)

**famsize** – dimensiunea familiei (‘LE3’=mai mică sau egală cu 3 membri sau ‘GT3’=mai mare de trei membri)

**Pstatus** – statusul părinților (‘T’=trăiesc impreuna sau ‘A’=divorțați)

**Medu** – educația mamei (0=fără, 1=educație primară (patru clase), 2=între 5 și 9 clase, 3=studii liceale sau 4=educație superioară)

**Fedu** - educația tatălui (0=fără, 1=educație primară (patru clase), 2=între 5 și 9 clase, 3=studii liceale sau 4=educație superioară)  
**Mjob** – locul de muncă al mamei (‘teacher’=profesor, ‘health’=în domeniul sănătății, ‘services’=în domeniul serviciilor, ‘at\_home’=casnică sau ‘other’=alte domenii de activitate)

**Fjob** - locul de muncă al tatălui (‘teacher’=profesor, ‘health’=în domeniul sănătății, ‘services’=în domeniul serviciilor, ‘at\_home’=casnic sau ‘other’=alte domenii de activitate)

**reason** – motivul pentru care au ales elevii școala unde studiază (‘home’=aproape de reședință, ‘reputation’=reputația școlii, ‘course’=în funcție de cursurile oferite sau ‘other’=alt motiv)

**guardian** – tutore legal/persoana care se ocupă de educația copilului (‘mother’=mama, ‘father’=tatăl sau ‘other’=alta persoana)

**traveltime** – distanța până la școală (1=mai puțin de 15 min., 2= între 15 și 30 min., 3= între 30 min. și 1 oră, sau 4=mai mult de 1 oră)

**studytime** – numărul de ore săptămânale dedicate învățatului (1= mai puțin de 2 ore, 2= între 2 și 5 ore, 3=între 5 și 10 ore, sau 4 = mai mult de 10 ore)

**failures** – numărul de corigențe din trecut (între 1 și 4)

**schoolsup** - sprijin educațional (yes=da sau no=nu)

**famsup** – ajutor educațional din partea familiei (yes=da sau no=nu)

**paid** – clase suplimentare plătite pentru limba pentru portugheză (yes=da sau no=nu)

**activities** – activități extra curiculare (yes=da sau no=nu)

**nursery** – a participat la grădinița (yes=da sau no=nu)

**higher** – dorește sa își continue studiile (yes=da sau no=nu)

**internet** – accesul la internet in rezedința (yes=da sau no=nu)

**romantic** – intr-o relație (yes=da sau no=nu)

**famrel** – calitatea relație dintre membrii familiei (între 1=foarte prost și 5=foarte bun)

**freetime** – timp liber după școală (între 1=foarte puțin și 5=foarte mult)

**goout** – ieșiri în oraș cu prietenii (între 1=foarte puțin și 5=foarte mult)

**Dalc** – consumul de alcool în timpul săptămânii (între 1=foarte puțin și 5=foarte mult)

**Walc** – consumul de alcool în timpul weekend-ului (între 1=foarte puțin și 5=foarte mult)

**health** – starea curentă de sănătate (între 1=foarte rău și 5=foarte bine)

**absences** – numărul de absențe (între 0 și 93)

**G1** – nota aferentă primului semestru (între 0 și 20)

**G2** – nota pe al doilea semestru (între 0 și 20)

**G3** – nota finală (între 0 și 20)

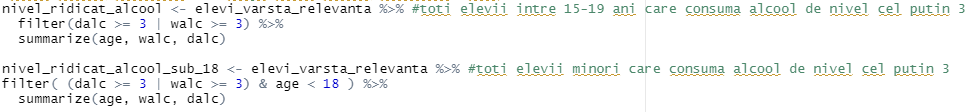
Considerăm că setul de date este relevant pentru că avem un număr mare de tineri care provin din medii diferite, avem foarte mulți factori care credem că au legătură cu performanțele lor academice și foarte mulți factori care le descriu viața într-un mod cât se poate de real și complet.

La partea de curățare a datelor nu am avut foarte mult de muncă, dar am executat câteva operațiuni, pentru a ne asigura că lucrăm cu un set de date corect și relevant. Astfel, am folosit librăria janitor, cu funcțiile pre-definite aferente: clean\_names (pentru a curăța denumirile tabelelor), tabyl (pentru a vedea dacă avem valori/câmpuri nule și anumite distribuții), remove\_empty – pentru a șterge orice eventuală coloană/rând complet nul. De asemenea, am folosit data.frame pentru a compara fostul set de date cu noul set de date, post-curățare. În ultimă instanță, din totalul de 370 de tineri de la școala gimnazială (clasele 5-9), i-am eliminat pe cei doi elevi cu vărstele de 20, respectiv 22 de ani. Motivul principal a fost faptul că reprezentau extremele din eșantionul nostru, rezultatele riscând să fie alterate într-un mod nedezirabil, pierzându-și din relevanță și acuratețe.

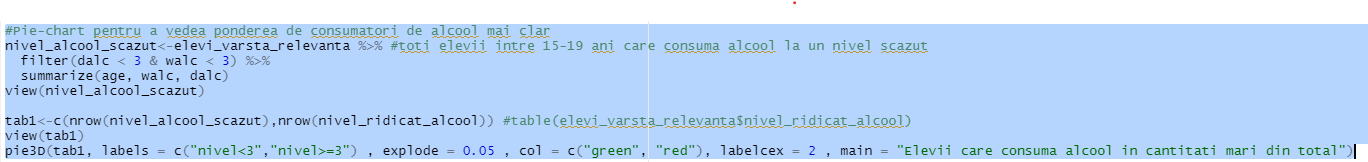
**Analiza datelor**

În analiza noastră am început cu prima întrebare, respectiv încercarea predicției succesului firmelor distribuitoare sau producătoare de alcool în viitorul îndepărtat. Pentru a putea face acest lucru posibil, pornim de la ipoteza nulă:

H0: Succesul firmelor distribuitoare sau producătoare de alcool nu depinde în nicio măsură de vârsta și nivelul de alcool consumat de către elevi.

Pentru a vedea cum stau lucrurile, am folosit funcția tabyl, în care am legat vârsta de nivelul de alcool consumat, atât în timpul săptămânii, cât și în weekend. În lumina unei analize relevante, am folosit un filtru pentru a selecta doar nivelurile ridicate de consum de alcool, respectiv de la 3 în sus (din maximul de 5), fie în weekend, fie în timpul săptămânii, fie pe parcursul întregii săptămâni. 

Cu ajutorul unui calculul simplu, constatăm că 40% dintre toți elevii (150 din 382) consumă alcool la un nivel ridicat. De asemenea, un fapt mai îngrijorător este că din acești 40%, 30% (111) sunt minori.



Chart

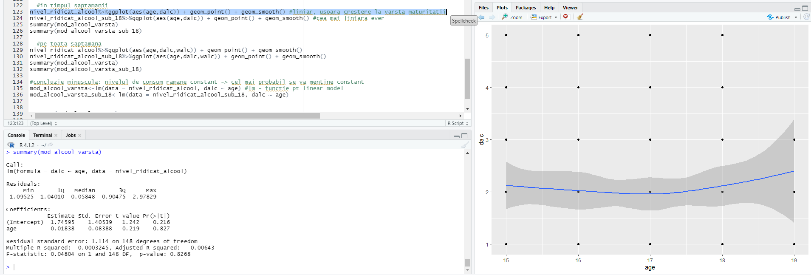
Description automatically generated

Graphical user interface, text, application

Description automatically generated

Am trasat și distribuția punctelor (cu funcția geom\_point()), în tandem cu trendul presupus liniar pe care îl urmează (cu geom\_smooth()).

În timpul săptămânii:



În weekend:



În continuare, pentru a obține indici consacrați în sprijinul analizei noastre am realizat grafice cu predicțiile pe viitor (cu add\_predictions), plecând de la ce avem în prezent:



În urma vizualizării graficelor, putem observa că tendința consumului de alcool în cantități medii spre mari, în rândul minorilor, este de ușoară scădere; antiteza apare în momentul în care vedem că tendința elevilor, pe toate vârstele, este de ușoară creștere. Explicația pe care am găsit-o este de ordin psihologic. Odată cu intrarea la liceu, apar influențe din partea colegilor, mai ales din clasele mai mari. Astfel, la 15 ani, o parte dintre copii încep să consume alcool. După ce trend-ul începe să se disipe, nemaifiind atât de în vogă, observăm o ușoară descreștere; urmează apoi o creștere, pe care o punem pe seama împlinirii vârstei de maturitate, respectiv 18 ani.

De asemenea, am făcut diverse modele (cu funcția glm), iar ca tip de regresie, am folosit regresia Poisson, deoarece avem date de numărare (vârsta de la 15 la 19, nivelul de alcool consumat de la 1 la 5):

Graphical user interface, text, application, email

Description automatically generated



Table

Description automatically generated

Am folosit și funcția tab\_model pentru a afla R2 – la noi se numește Nagelkerke, dar este adaptarea de la R2 de la funcția liniară. Ne arată că modelul nu este unul foarte bun.

Așa cum vedeți mai sus, în cazul consumului de alcool în weekend, descoperim o mică problemă: p-value pentru intercept este aproape 0.5, deci nu îl vom lua în considerare. Cu toate acestea, în lumina unui răspuns mult așteptat la întrebarea de cercetare, observăm că p-value, pentru variabila vârstă, este 0.0194, ceea ce arată o conexiune suficient de puternică între cele 2 variabile; adăugăm că panta este +0.07067 și eroarea standard este de 0.03024; folosind formula βi ∈ [βˆi − 2SE(βˆi ), βˆi + 2SE(βˆi )] pentru CI, obținem intervalul [0.0102, 0.13115], deci înaintarea în vârstă determină un consum mai pronunțat de alcool pe timpul weekendului, chiar dacă în mică măsură.

Text

Description automatically generated



Table

Description automatically generated

În partea cealaltă, când discutăm despre consumul de alcool în timpul săptămânii, situația stă altfel. P-values depășesc pragul consacrat de 0.05, ceea ce arată o conexiune inexistentă între cele 2 variabile – lucru arătat atât de summary, cât și de tab\_model.

Per total avem următorul caz:



Am folosit funcția rowMeans pentru a face media consumului de alcool per elev, pentru a obține o valoarea pentru toată săptămâna; parte din valori sunt de tipul float, așa că am creat un set de date cu valorile dublate, pentru ca regresia poisson să le poată procesa. Astfel, vom obține, la rulare, următoarele informații:

Text

Description automatically generated



Table

Description automatically generated

Aici, p-value pentru vârstă are o valoare de sub 0.01, ceea ce sugerează o relație foarte puternică între cele 2 variabile. Panta este +0.06226, iar eroarea 0.02359, de unde putem deduce că intervalul este [0.015, 0.11] adică înaintarea în vârstă determină un consum mai accentuat de alcool, chiar dacă la un nivel foarte mic (dacă un om împlinește vârsta de 18 ani, conform intervalului de încredere înseamnă că nivelul îi crește, să zicem, de la 2 la 2.02-2.11).

Bonus: Dacă luăm în calcul extrema stângă a intervalului, nivelul de consum după 50 de ani va ajunge de la 2 la 2.75 – practic, o diferență insesizabilă pentru un om oarecare, pe o perioadă așa de îndelungată. Dacă luăm în calcul și alți factori necuprinși în analiză sau seturi de date pe care nu le avem în posesie (de la facultate, modificări majore apărute în viață), nivelul ar putea stagna sau chiar scădea în timp; un motiv bun în sprijinul acestei idei îl constituie formarea unei familii și/sau intrarea în câmpul muncii.

Observație: Dacă ne uităm la toate valorile R2, observăm că cel mai bun model este cel cu consumul de alcool pe toată săptămâna.

Astfel, concluzia cercetării până în acest punct este că ipoteza nulă de la care am plecat este respinsă, adică vârsta determină nivelul consumului de alcool. Legătura este puternică, dar cu o influență redusă. În această ordine de idei, am putea presupune că firmele vor avea succes peste 10-20 de ani în distribuția și/sau prelucrarea alcoolului, strict pe baza analizei, adresându-se generației actuale.

Presupunerea ar fi putut deveni predicție dacă aveam și un set de date (pentru același eșantion) din perioada facultății și, eventual, a primilor 5 ani de după.

Cu toate acestea, ca recomandare, administratorii s-ar putea concentra asupra grupului țintă de acum, menținând în stoc cele mai vândute băuturi ale tinerilor, cu speranța că tendința pe care am identificat-o noi își va păstra cursul și pe perioada următorilor ani.

Metoda II: Arbori

Am ales arborii deoarece rezultatele sunt ușor de interpretat, mai ales pentru persoane care nu sunt din domeniu, lucru excelent în cazul nostru (șansele ca proprietarii de magazine să știe statistică sunt foarte mici).

Vedem, în primul rând, ce date numerice avem (19 din 36 de coloane).

Chart

Description automatically generated

Împărțim apoi setul de date în date de antrenament, respectiv de test, pentru a face calculele și predicțiile.

A picture containing text

Description automatically generated

Verificăm apoi arborii pentru influența vârstei asupra consumului de alcool, în cele 2 cazuri:

1.În weekend:

Text

Description automatically generated with low confidence

Diagram

Description automatically generated with low confidence

A picture containing graphical user interface

Description automatically generated

2.În timpul săptămânii

Text

Description automatically generated with low confidence

Graphical user interface, application

Description automatically generated

A picture containing graphical user interface

Description automatically generated

Erorile sunt destul de mari aici, raportat la nivelul de consum.

**Întrebarea 2: Influențează consumul alcoolului performanțele academice**?

Text, letter

Description automatically generated

Table

Description automatically generated

Text, letter

Description automatically generated

Table

Description automatically generated

Text, letter

Description automatically generated

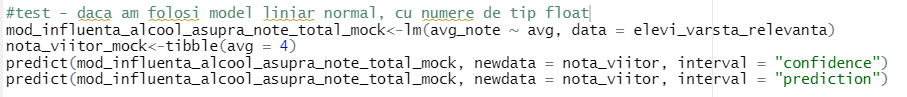
Table

Description automatically generated

Cel mai bun model, conform R2, este al doilea, respectiv cel cu limba portugheză. Putem vedea acest lucru la p-value, când comparăm cu ultimul model, cel per total.

Ne vom uita, în primul rând, la p-value; în cazul matematicii, valoarea fiind de 0.341, realizăm că nu există o relație între aceasta și consumul de alcool; în cazul limbii portugheze, lucrurile stau complet pe dos - p-value este <0.001, adică legătura dintre cele două este foarte puternică. Același lucru este și la calculul per total, unde p-value este tot <0.001. Fiind cazul cel mai reprezentativ, ne uităm la pantă, care ia valoarea -0.021, și la eroare, care ia valoarea 0.00565, de unde și rezultă intervalul [-0.0322, -0.01]; se poate vedea clar că un consum de alcool crescut scade performanțele academice ale elevilor, așa cum era și de așteptat. Consumul de alcool vine la pachet cu mai multe riscuri decât beneficii, așa cum putem vedea atât din analiza desfășurată de noi, cât și din articolele științifice redactate de cercetători.

**Test:**

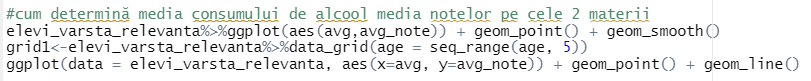


Text

Description automatically generated

...până la 368

Mai jos descoperim cum influențează consumul de alcool per total media dintre nota la matematică și nota la portugheză, precum și distribuția.



Chart, line chart

Description automatically generated

Predicții: Graphical user interface, website

Description automatically generated with medium confidence

Observăm că un consum de nivel 5 (10 e 5, de fapt – am folosit artificiul la calcul) aduce cu sine note foarte mici.

Metoda II: Arbori de decizie

1. Matematică

A picture containing text

Description automatically generated

Diagram

Description automatically generated



Observăm, în cazul matematicii, că elevii care consumă alcool în timpul săptămânii la un nivel redus (automat și în weekend) au cea mai mare medie a notelor, respectiv 11 (și, ce e drept, majoritatea nici nu prea consumă – 68%). De asemenea, elevii cu un consum ridicat de alcool în weekend (automat și în timpul săptămânii) au media notelor 7.4 – o diferență ca de la cer la pământ.

1. Portugheză

A picture containing text

Description automatically generated

Diagram

Description automatically generated

A picture containing text

Description automatically generated

La portugheză observăm că arborele nu ia în considerare nivelul de alcool consumat în weekend, așadar este mult mai relevant nivelul de alcool consumat din timpul săptămânii. Cei care beau de nivel peste 4 au media 9, cei peste 3 media 11, iar cei care consuma sub nivelul 3 au media 13!

Eroarea medie pătratică (RMSE) este aproape 3, ceea ce în contextul nostru este o eroare mare! În sensul că dacă noi prezicem că nota unui elev este 11, poate fi între 8-14.

Întrebarea 3: Influențează notele consumul de alcool?

Dorim să vedem dacă emoțiile asociate primirii unei note mari sau mici determină un elev să consume alcool.

Text, letter

Description automatically generated

Table

Description automatically generated

P-value este 0.006, arătând o conexiune puternică între variabile; panta este -0.01111 și eroarea 0.00404, de unde deducem că intervalul de încredere este [-0.02, -0.003]. Concludem că da, notele obținute influențează consumul de alcool, în mod invers proporțional și foarte fin. Cu cât nota este mai mare, cu atât dorința elevului de a consuma alcool se diminuează. Explicația oferită de noi vine pe parte psihologică - sentimentul de bucurie, încrederea și sima de sine sporită scad doza de teribilism și nevoia de atenție. Așadar, performanțele academice crescânde formează, pe lângă cunoștințele profesionale, și un caracter dezvoltat armonios.

Predicții: Text

Description automatically generated with medium confidence

Putem observa că dacă un elev are media notelor finale 20, adică 20 la mate și 20 la portugheză (maximul posibil), el va avea un nivel de consum, în medie, de 1.56, care este foarte mic și benefic pentru elev.

Arbori:

Text

Description automatically generated

1. Nota la matematică influențează consumul de alcool din weekend

Diagram, timeline

Description automatically generated

Graphical user interface, text

Description automatically generated with medium confidence

1. Nota la matematică influențează consumul de alcool din timpul săptămânii

Diagram, timeline

Description automatically generated

Graphical user interface, text

Description automatically generated with medium confidence

1. Nota la portugheză influențează consumul de alcool din weekend

Diagram

Description automatically generated with medium confidence

Aici e cât se poate de simplu: Dacă nota pe an la portugheză este mai mică de 13, atunci nivelul de alcool este, în medie, 2.6, versus 2.2 în cazul celor cu note peste. Împărțirea este aproape 50-50.

Graphical user interface, text

Description automatically generated with medium confidence

1. Nota la portugheză influențează consumul de alcool din timpul săptămânii

Diagram

Description automatically generated

Graphical user interface, text

Description automatically generated with medium confidence

Vă mulțumim pentru atenția acordată!