Raport Analiza Inferentiala

Contents

[1 Introducere 2](#_Toc59459852)

[2 Configurarea si premizele analizei 3](#_Toc59459853)

[2.1 Elemente de configurare a modulului de analiza inferentiala 3](#_Toc59459854)

[2.2 Sumarul datelor analizate 5](#_Toc59459855)

[3 Analiza statistica a datelor 7](#_Toc59459856)

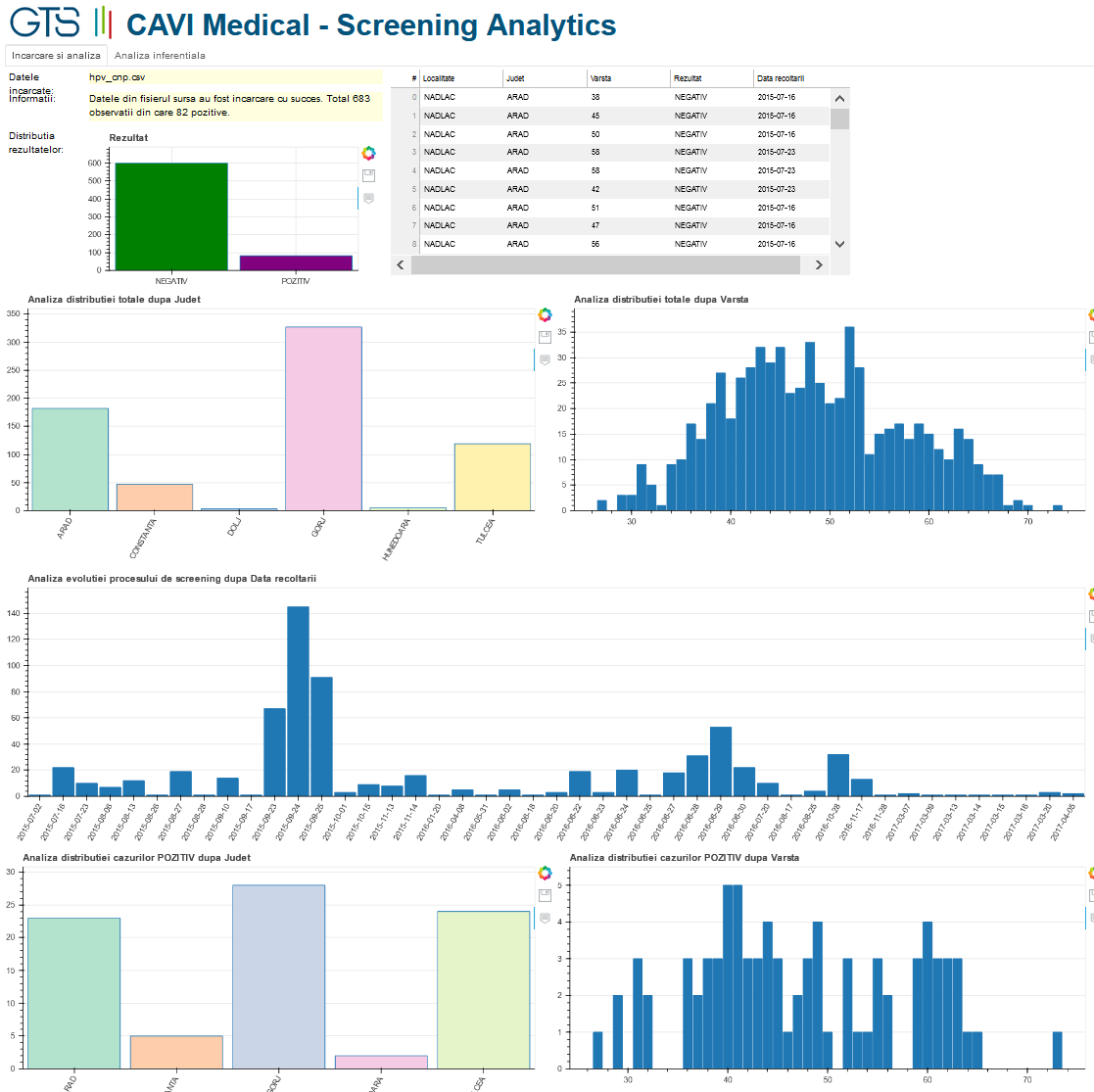
[3.1 Istoricul colectarii 7](#_Toc59459857)

[3.2 Analiza distributiilor 8](#_Toc59459858)

[4 Analiza inferentiala a datelor 12](#_Toc59459859)

# Introducere

In cadrul acestui raport sunt prezentate principalele rezultate ale analizei inferentiale automatizate realizate prin sistemul CAVI Medical Screening Analytics. Modulele de analiza inferentiala permit incarcarea de date pe formate variabile culese in campaniile de screening medical si analiza acestora atat statistica cat si inferentiala.



# Configurarea si premizele analizei

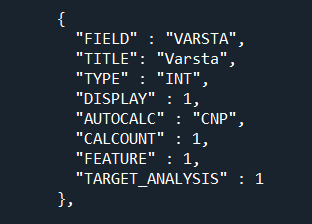
## Elemente de configurare a modulului de analiza inferentiala



Figure - Structura completa de configurare a sistemului in baza unui fisiere text de tip JSON

Dupa cum se poate observa din imaginea de mai sus modulul de analiza inferentiala poate fi configurat pe aproape orice surse de date prin asocierea campurilor din fisierul CSV-ul sursa cu diverse tipuri analiza statistica a datelor. Din perspectiva tipurilor de date mentionam urmatoarele categorii principale:

* Campuri de tip numeric (INT) in baza carora se pot calcula si analiza informatiile prin afisari de coloane sau distributii. In cazul campului prezentat mai jos se face automat generarea valorii varstei in baza coloanei CNP din date surse care ulterior din motive de confidentializare este total eliminate din analiza. Concret in aceasta situatia campul VARSTA este o auto-calculare bazata pe analiza datelor in format-ul specific coduluio numeric personal.



* Campuri de tip date categorice care se pre-proceseaza pentru eliminarea erorilor de introducere si ulterior pot utiliza pentru analiza statistica a distributiilor. Mai jos avem doua exemple: in primul exemplu avem un cam care se afiseaza in tabelele de date fara insa a fi utilizat pentru analize statistice iar in al doilea caz un camp categoric care este folosit inclusive pentru analize

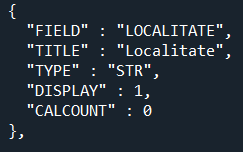


Figure - Camp doar pentru afisare

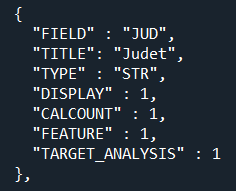


Figure - Camp pentru care se calculeaza si afiseaza analize statistice

* Campurile de tip data configurabile conform urmatoarei imagini sunt utile pentru analiza evenimentelor de recolatare in timp

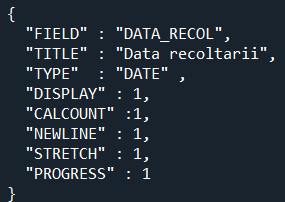


Figure - Configurarea campurilor de tip data (data recoltarii)

* Nu in ultimul rand avem campurile de tip rezultat care sunt marcate cu atributul TARGET conform imagii de mai jos

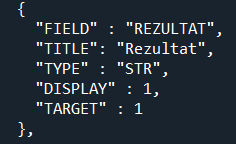


Figure - Definirea campului din care se extrage valoarea tinta.

* In configurarea modulului inferential exista si alte elemente de definire a meta-datelor dintre care cel mai important este identificarea in sursa de date a categoriei de observatii care reprezinta cazuri positive. Pentru acest scop la nivelul configurarii se defineste variabila TARGETED\_VALUE care va trebui sa aiba valoara regasibila in sursa de date. In cazul nostrum sursa de date CSV contine valori “POZITIV” pentru cazurile cu leziuni confirmate in coloanal REZULTAT.



Figure - Definirea valorii pentru cazurile positive

## Sumarul datelor analizate

In vederea livrarii prezentului raport de analiza inferentiala au fost incarcate date rezultate din campanii de screening de col uterin desfasurate in perioada 2015-2017 in cadrul a mai multe campanii/proiecte. S-au utilizat doua baze de date dupa cum urmeaza:

* Testare Babes-Papanicolaus pe un numar de 2592 de cazuri
* Testare Human Papiloma Virus pe un numar de 683 de cazuri

Cele doua surse de date au avut mici diferente din punct de vedere structural si mai ales din punct de vedere al rezultatelor dupa cum se poate vedea din graficele celor doua distributii ale rezultatelor de mai jos.

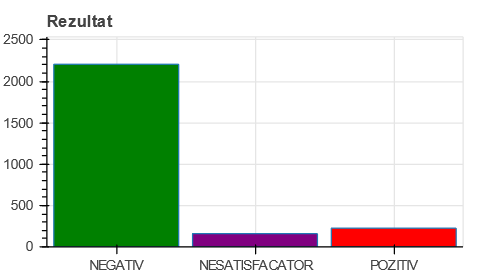


Figure - Distributia observatiile pentru datele aferente testelor Babes Papanicolaus

.

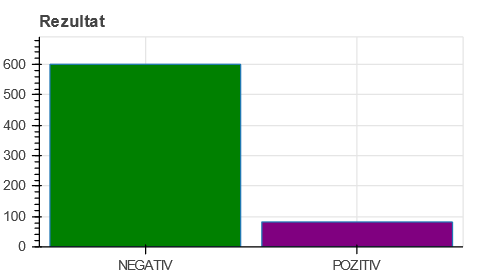


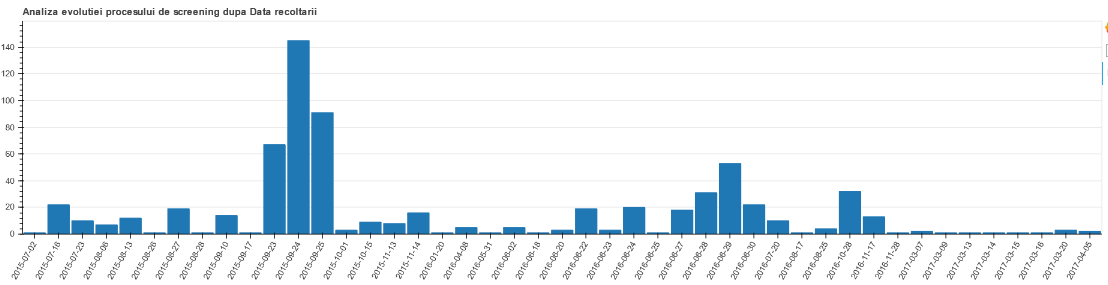
Figure - Distributie rezultatelor pentru testele Human Papiloma Virus

# Analiza statistica a datelor

Sub-modulul (tab-ul) de analiza grafica statistica din cadrul modulului de analiza inferentiala se bazeaza pe configuratia definite conform descrierii din capitolul anterior. Astfel, este posibila analiza a multiple variabile regasibile in sursa de date sau auto-calculabile.

## Istoricul colectarii

Modulul de analiza permite afisarea istoricului colecatarii si evaluarii subiectilor conform graficului prezentat mai jos



In acest grafic se se poate observa dinamica realizarii de recoltarii de la prima data de inceput a proiectului de screening pana la ultima recoltare.

De mentionat este ca faptul ca toate aceste imagini sunt obtenabile si exportabile direct din modulul de analiza inferentiala prin uneltele atasate fiecarui grafic conform imaginii de mai jos

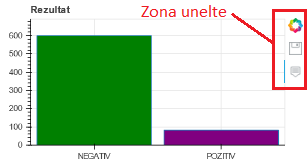


Figure - Exportarea de grafice statistice din modulul de analiza inferentiala

## Analiza distributiilor

Toate variabilele care au fost marcate ca fiind CALCOUNT = 1 in configurare vor fi analizate din punct de vedere al distributiilor. In cele ce urmeaza vom prezenta rezultatele pe cele doua surse de date descrise anterior.

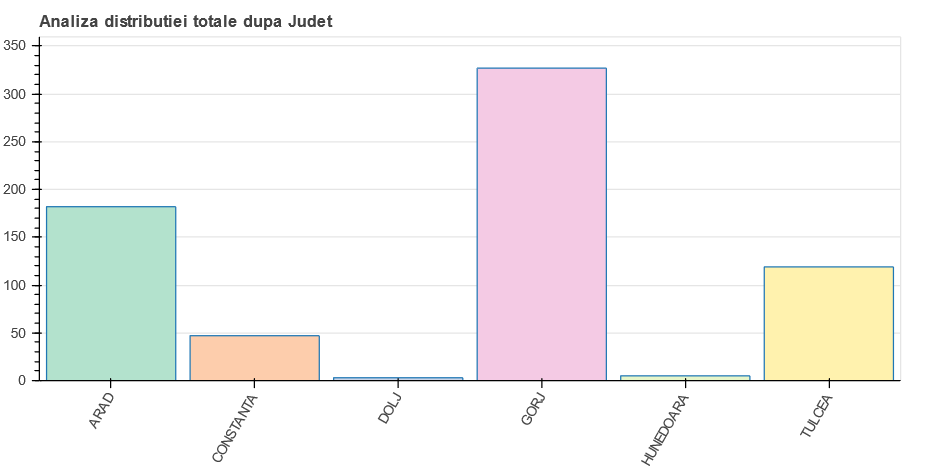


Figure - Analiza distributiei de judete implicate in proiectul de screening prin testare HPV in perioada 2015-2017

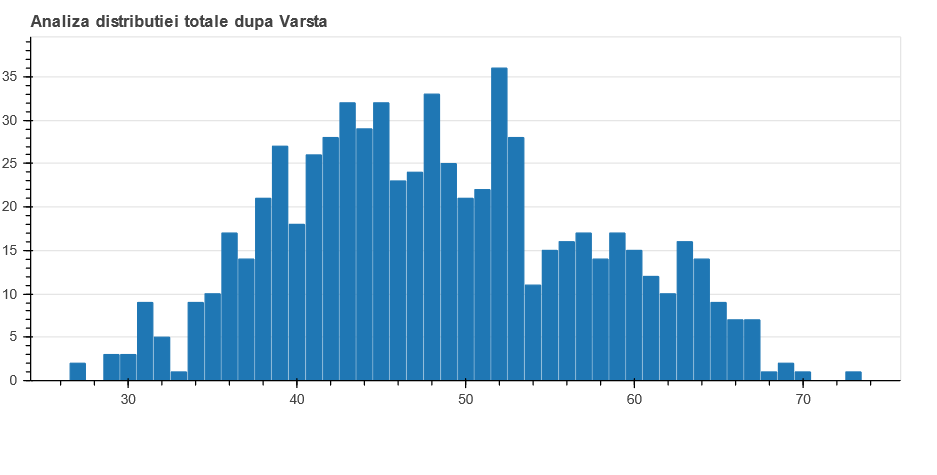


Figure - Analiza distributiei varstelor pentru toata populatia de 683 de femei testate HPV

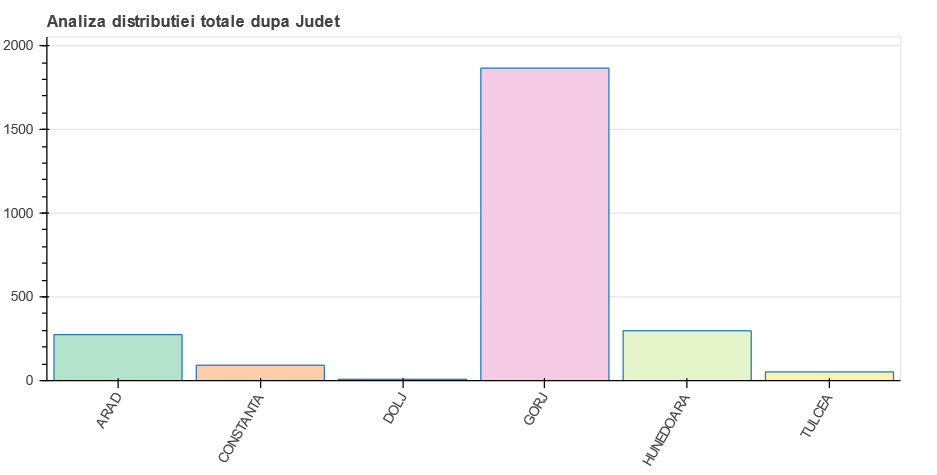


Figure - Analiza numarului de teste per judet pentru proiectul bazat pe testare Babes Papanicolau

Analiza distributiilor facuta la nivelul intregii populatii prezentata mai sus pentru ambele proiecte de screening analizate este automat replicata de catre modulul de analiza inferentiala filtrand doar cazurile pozitive dupa cum urmeaza sa prezentam in urmatoarele patru grafice

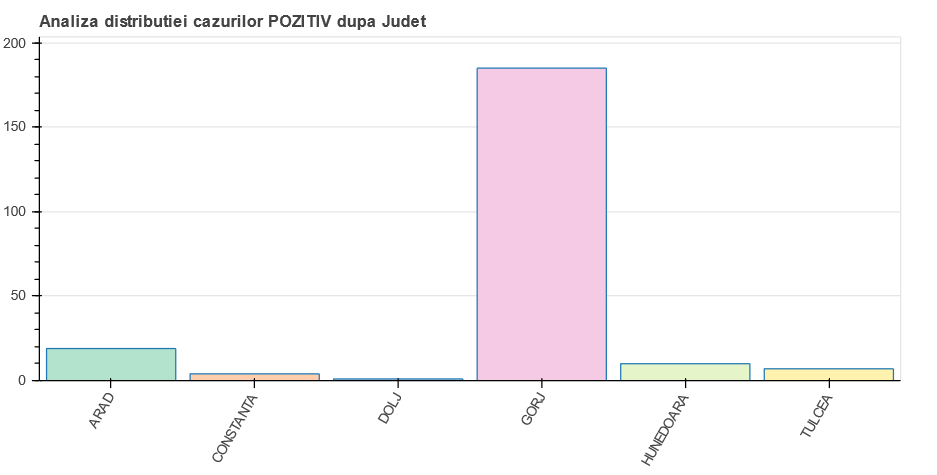


Figure - Analiza distributiei pe judete a cazurilor pozitive pentru cele 226 de teste pozitive Babes Papanicolau 2015-2016

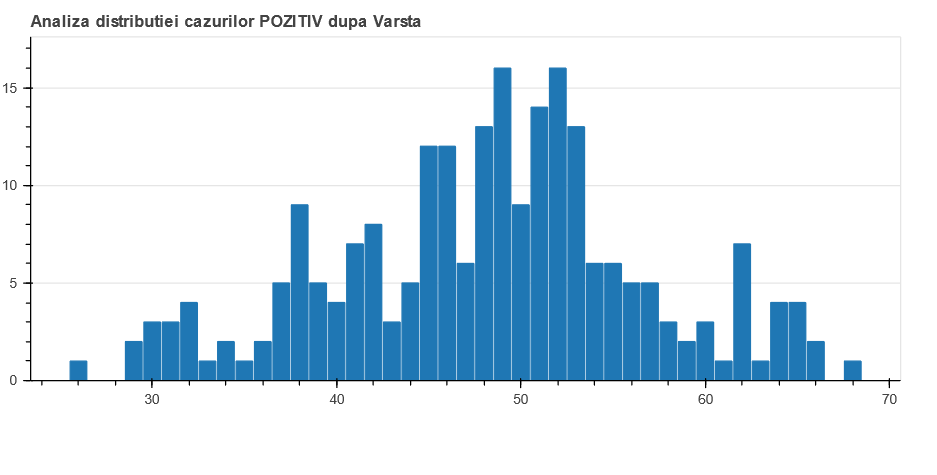


Figure - Analiza celor 226 cazuri pozitive din cele 2592 teste Babes Papanicolu, din perspectiva distributiei varstelor, teste realizate intre 2015 si 2016

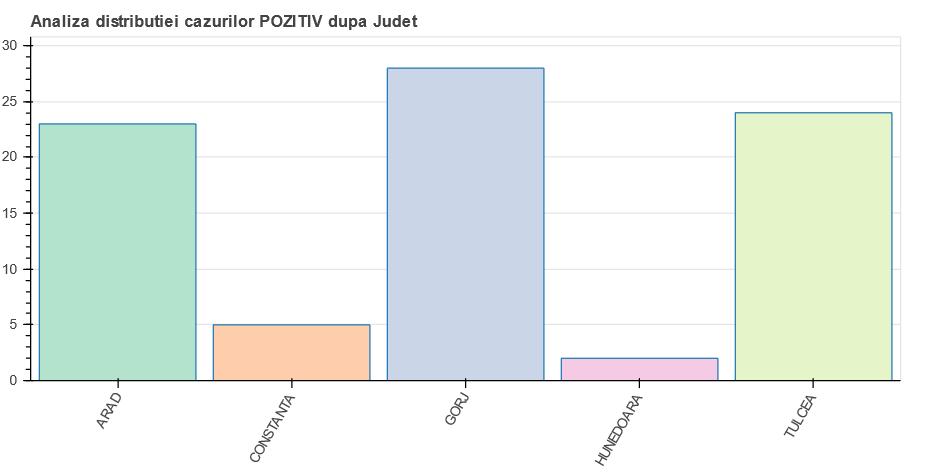


Figure - Analiza distributiei cazurilor pozitive HPV la nivel de judet pentru cele 82 de cazuri pozitive din 683 de teste HPV realizate intre 2015-2017

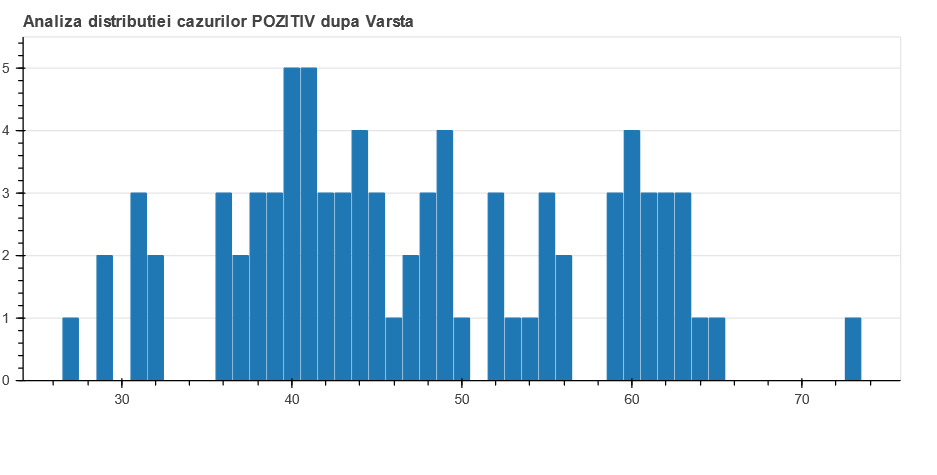


Figure - Analiza distributiei dupa varsta subiectului pentru cele 82 de cazuri pozitive HPV din cele 683 de teste realizate in perioada 2015-2017

# Analiza inferentiala a datelor

Pentru analiza inferentiala s-a optat realizarea unui model inferential antrenat prin invatare automata care a determinat corelatia intre diversele variabile analizate si probabilitatea crescuta de a exita o leziune.

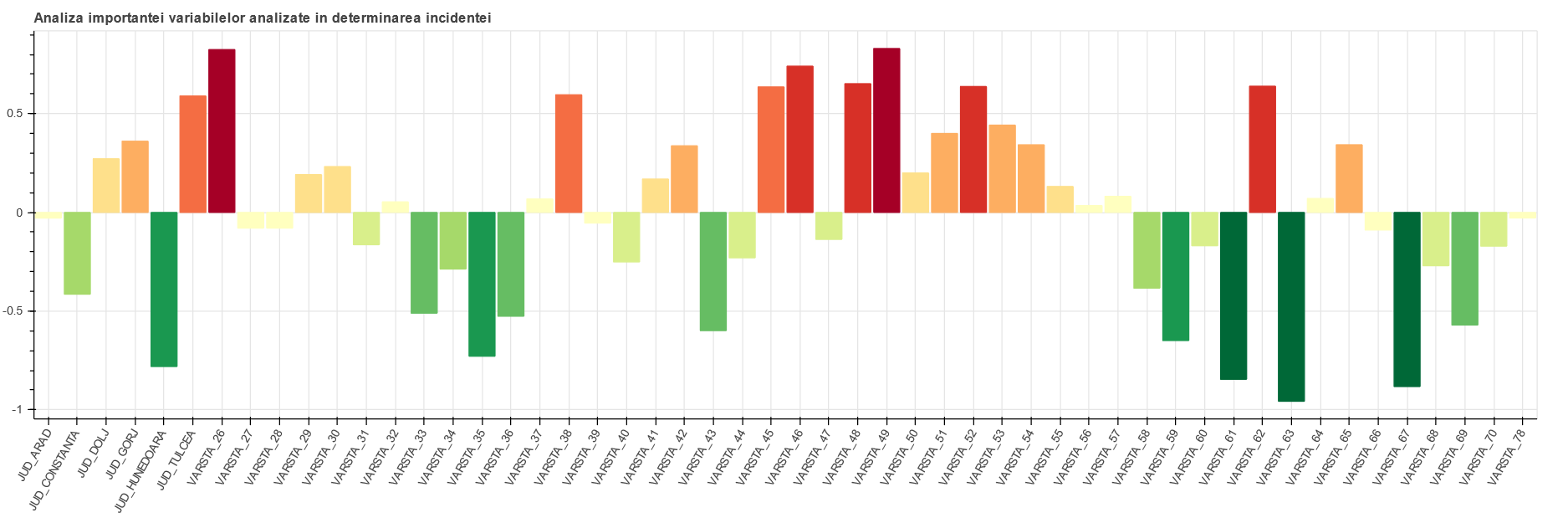


Figure - Graficul importantei si corelatiei diverselor variabile in determinarea probabilitatii existentie unei leziuni

In graficul prezentat in Figure 17 sunt enumerate principalele variabile analizate iar codurile de culoara reprezinta influenta pozitiva (rosu) versus influenta negativa (verde) a respectivei variabile in nivelul de probabilitate a existentei unui caz pozitiv. Mai concret vom analiza sectiunea de inceput a aceste analize dupa cum urmeaza.

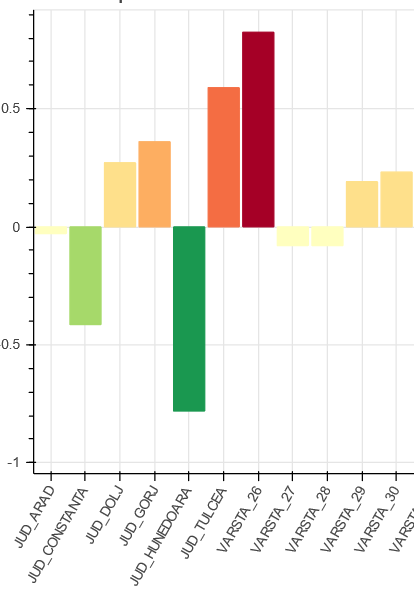


Figure - Analiza a 11 variabile si impactul acestora asupra nivelului de probabilitate de existanta a unui caz pozitiv

Din graficul de mai sus se observa ca, in baza datelor colectate pentru cele 226 cazuri pozitive din totalul de 2592 de cazuri avem influente puternic neative pentru cazurile din judetul Hunedoara – apartenenta in aceasta zona deteminand probabilitate mai mica de existenta a unei leziuni – fara de judetul Tulcea unde apartenenta in acest al doilea judet determina o crestere majora a riscului de existenta a unei leziuni. Deasemenea se observa ca varsta de 26 de ani implica un risc mult mai mare decat 27-29.

Similar putem analiza acelasi grafic de analiza inferentiala si pentru datele recolatate in baza testarii HPV.

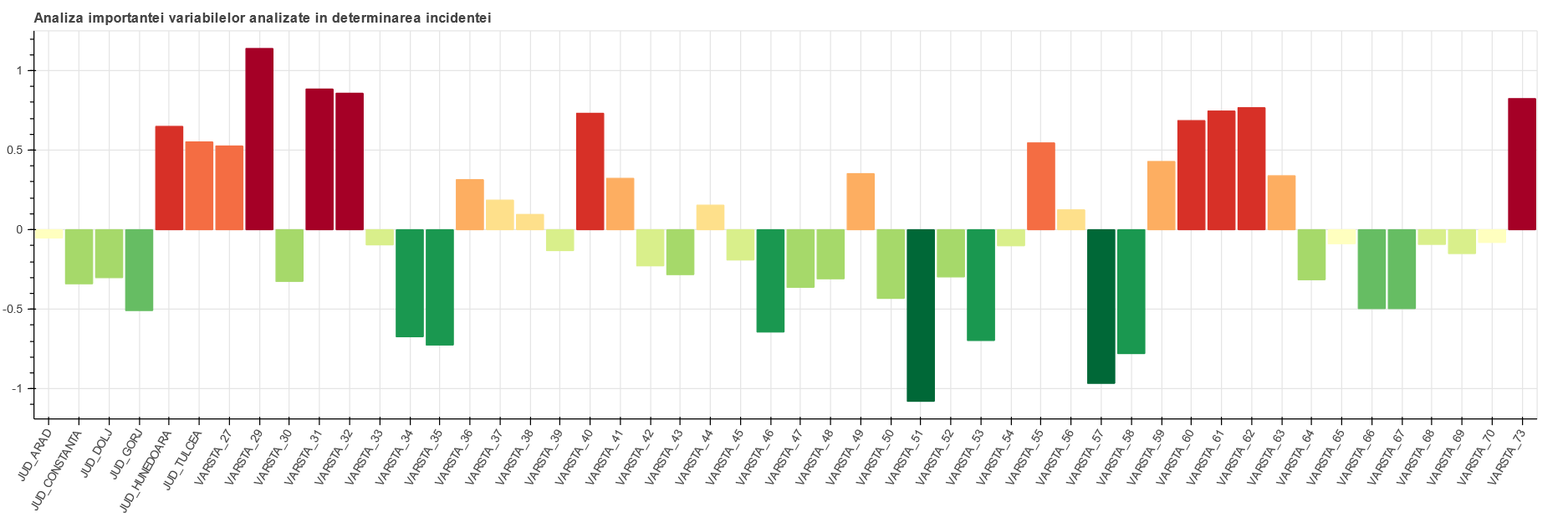


Figure - Analiza inferentiala a impactului diverselor variabile pentru nivelul de risc al infectiei HPV