

Dataset Exploration Report

1. Dataset Overview

- Dataset Name: **Oral Cancer Prediction Dataset**
 - Source: [Link](#)
 - Number of Records (Rows): 84922
 - Number of Features (Columns): 25
-

2. Data Dictionary (Summary of Features)

Feature name	Data type	Description
Age	Numerical	Patient's age in years
Gender	Categorical	Patient's gender (Male/Female)
Tobacco Use	Categorical	Whether the patient uses tobacco (Yes/No)
Alcohol Consumption	Categorical	Whether the patient consumes alcohol (Yes/No)
HPV Infection	Categorical	Presence of Human Papillomavirus infection (Yes/No)
Betel Quid Use	Categorical	Whether the patient chews betel quid (Yes/No)
Poor Oral Hygiene	Categorical	Self-reported poor oral hygiene (Yes/No)
Oral Lesions	Categorical	Presence of oral lesions (Yes/No)
White or Red Patches in Mouth	Categorical	Presence of abnormal patches in the mouth (Yes/No)
Difficulty Swallowing	Categorical	Difficulty in swallowing food or liquids (Yes/No)
Tumor Size (cm)	Numerical	Size of tumor in centimeters (continuous)
Cancer Stage	Numerical	Stage of oral cancer (0 - 4)

Early Diagnosis	Categorical	Whether the cancer was diagnosed early (Yes/No)
Oral Cancer (Diagnosis)	Categorical	Target variable: whether the patient is diagnosed with oral cancer (Yes/No)

3. Summary Statistics

- Numerical Features:

	Age	Tumor Size (cm)	Cancer Stage	Survival Rate (5-Year, %)	Cost of Treatment (USD)	Economic Burden (Lost Workdays per Year)
count	84922	84922	84922	84922	84922	84922
mean	54.509	1.747	1.118	79.503	39109.88	52.028
median	55	0	0	100	0	0
std	10.014	2.028	1.34	26.483	44710.68	60.389
min	15	0	0	10	0	0
max	101	5.999	4	100	159988.00	179

- Categorical Features:

Feature	Top Category	Frequency	unique
Country	India	8079	17
Gender	Male	60351	2
Tobacco Use	Yes	67989	2
Alcohol Consumption	Yes	50939	2
HPV Infection	No	59525	2
Betel Quid Use	No	60299	2
Chronic Sun Exposure	No	68057	2
Poor Oral Hygiene	No	42627	2
Diet (Fruits & Vegetables Intake)	Moderate	33914	3
Family History of Cancer	No	76508	2
Compromised Immune System	No	76368	2
Oral Lesions	No	59419	2
Unexplained Bleeding	No	67891	2
Difficulty Swallowing	No	63709	2
White or Red Patches in Mouth	No	55103	2
Treatment Type	No Treatment	51016	5
Early Diagnosis	Yes	46508	2
Oral Cancer (Diagnosis)	No	42573	2