* **Project Documentation: Dentistry Dataset Analysis**

**Table of Contents**

1.Introduction

2.Dataset Description

3.Data Loading

4.Data Preprocessing

5.Analysis

6.Conclusion

7.References

**1. Introduction**

This project aims to analyse the Dentistry Dataset to extract insights related to dental health and patient demographics. The analysis will help in understanding trends, patient needs, and potential areas for improvement in dental care.

**2. Dataset Description**

Dataset Name: Dentistry Dataset

File Format: CSV

# Get the number of rows and columns

num\_rows, num\_columns = df.shape

df.shape

**4. Data Loading**

To load the dataset, use the following code:-

**import pandas as pd**

**import Numpy as np**

**import matplotlib.pyplot as plt**

**import seaborn as sns**

**from sklearn.preprocessing import LabelEncoder**

df = pd.read\_csv("Dentistry Dataset.csv")

This will read the dataset into a Pandas Data Frame named df.

**5. Data Preprocessing**

Before analysis, the dataset may require preprocessing. Common steps include:-

**# Handle missing values**

missing\_values = df.isnull().sum()

(missing\_values)

**6. Analysis**

Perform various analyses to derive insights. This may include:

**import pandas as pd**

**from sklearn.preprocessing import LabelEncoder**

df = pd.read\_csv("Dentistry Dataset.csv")

**# Create a label encoder**

label\_encoder = LabelEncoder()

**# Encode the 'Age' column**

df['Age'] = label\_encoder.fit\_transform(df['Age'])

# Access the 'Age' column

(df['Age'])

**# Split independent and dependent variables i.e. X and Y**

X = df.drop('Age', axis=1) # Independent variables

y = df['Gender'] # Dependent variable

# the shapes of X and y

("Shape of X:", X.shape)

("Shape of y:", y.shape):

**# Split into independent (X) and dependent (y) variables**

X = df.drop('Age', axis=1) # Drop the target variable column

y = df['Inter-canine Distance Intraoral'] # Assuming this is the dependent variable

# Display the normalized data

print(X\_normalized\_df.head())

**import pandas as pd**

**import matplotlib.pyplot as plt**

# Load the data from a CSV file

df = pd.read\_csv("Dentistry Dataset.csv")

plt.scatter(df['Age'], df['inter canine distance intraoral'])

plt.title('Scatter Plot of Age vs. Inter-canine Distance')

plt.xlabel('Age')

plt.ylabel('Inter-canine Distance Intraoral')

plt.show()

import pandas as pd

import matplotlib.pyplot as plt

# Load the data from a CSV file

df = pd.read\_csv("Dentistry Dataset.csv")

# Display the first few rows of the dataframe

("Original DataFrame:")

(df.head())

df.drop(columns=unwanted\_columns, inplace=True)

# Display the updated dataframe

print("\nUpdated DataFrame after dropping unwanted columns:")

print(df.head())

# scatter plot for the remaining important variables

plt.scatter(df['Age'], df['inter canine distance intraoral'])

plt.title('Scatter Plot of Age vs. Inter-canine Distance')

plt.xlabel('Age')

plt.ylabel('Inter-canine Distance Intraoral')

plt.show()

**Summary of Findings**

1. **Data Structure**: The Dentistry Dataset was successfully loaded and analysed, revealing a variety of features related to dental health and patient demographics. The analysis primarily focused on the relationship between age and inter-canine distance, which could provide insights into dental growth patterns.
2. **Age Encoding**: The 'Age' column was encoded using Label Encoder, transforming categorical age values into numerical format. This is useful for machine learning applications but may introduce ordinal implications that need to be addressed.
3. **Missing Values**: Initial checks for missing values were conducted, highlighting areas that may require attention to ensure data integrity.
4. **Scatter Plot Insights**: The scatter plot visualized the relationship between age and inter-canine distance. Observing trends or patterns in this plot can indicate how age influences dental measurements, which can inform dental practices.
5. **Data Cleanup**: Unwanted columns were dropped to streamline the analysis, allowing focus on relevant features that contribute meaningfully to understanding dental health.