WINE DATA ANALYSIS

HLD Document

DOCUMENT VERSION CONTROL

Date Issue	Version	Description	Author
07/04/2023	1	Initial HLD – V 1.0	SHRUTI
14/04/2023	2	Final HLD – V 2.0	SHRUTI

CONTENT

- DOCUMENT VERSION
- ABSTRACT
- INTRODUCTION
- WHY THIS HIGH-LEVEL DESIGN DOCUMENT?
- SCOPE
- GENERAL DESCRIPTION
- PROBLEM STATEMENT
- PRODUCT PERSPECTIVE
- PROPOSED SOLUTION
- TECHNICAL REQUIREMENTS
- TOOLS USED
- POWER BI INTRODUCTION
- DESIGN DETAILS
- OPTIMIZATION
- ARCHITECTURE DESCRIPTION
- POWER BI DASHBOARD INTRODUCTION
- WINE QUALITY DASHBOARD
- WINE QUALITY PREDICTOR DESCRIPTION

ABSTRACT

Wine is an alcoholic beverage made from fermented grapes or other fruits. The process of making wine involves crushing and fermenting the fruit, usually grapes, to produce a liquid that contains alcohol.

Wine is classified based on several factors, including the type of grape used, the region where the grapes were grown, and the production methods used. Some common types of wine include red wine, white wine, rosé wine, and sparkling wine. Wine can also be classified as dry, semi-dry, or sweet, depending on the residual sugar content.

The objective of the project is to build a predictor that will help to understand the wine quality and its data pattern. The goal is to model wine quality based on physicochemical tests

INTRODUCTION

WHY THIS HIGH-LEVEL DESIGN DOCUMENT?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

THE HLD WILL -

- PRESENT ALL THE DESIGN ASPECTS AND DEFINE THEM IN DETAIL
- DESCRIBE THE USER INTERFACE BEING IMPLEMENTED
- DESCRIBE THE HARDWARE AND SOFTWARE INTERFACE
- DESCRIBE THE PERFORMANCE REQUIREMENT
- LIST AND DESCRIBE THE NON-FUNCTIONAL ATTRIBUTES SECURITY

1. RELIABILITY
2. MAINTAINABILITY PORTABILITY
3. REUSABILITY
4. APPLICATION COMPATIBILITY RESOURCE UTILIZATION
5. SERVICEABILITY

SCOPE

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

GENERAL DESCRIPTION

PROBLEM STATEMENT

There are Red and White Vinho Verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests.

This dataset contains details of the alcohol, residual sugar, acidity, density etc. and about having quality of wine. Using the attributes in the dataset we can predict quality of wine and identify the factors that affect the quality of wine.

The main objective of the project includes:

- Analyzing raw dataset.
- Perform data cleaning and preprocessing operations on the raw data.
- Building visualizations on the cleaned features using a BI tool.
- Building Dashboard using Power BI
- Building Predictor using Streamlit

TECHNICAL REQUIREMENTS

The solution can be a cloud-based or application hosted on an internal server or even be hosted on a local machine. For accessing this application below are the minimum requirements:

- Good internet connection.
- Web Browser.

For training model, the system requirements are as follows:

- +4 GB RAM preferred
- Operation System: Windows, Linux, Mac
- Power Bi / Jupyter Notebook / Google Collab

TOOLS USED

Python Libraries such as NumPy, Pandas, Matplotlib, Seaborn, Streamlit Business Intelligence tools such as Power BI, Collab / Jupyter Notebook and Python Programming Language are used to build the whole framework.

POWER BI

Power BI is a suite of business analytics tools developed by Microsoft that allows users to analyze data and share insights. It includes a collection of software services, apps, and connectors that work together to turn data into interactive visualizations and business intelligence reports.

With Power BI, users can connect to a wide variety of data sources, including Excel spreadsheets, cloud-based and on-premises data sources, and even streaming data. Once the data is connected, users can create custom reports, dashboards, and visualizations that can be easily shared with others in their organization.

Power BI also includes features such as natural language query, which allows users to ask questions of their data using plain English, and machine learning capabilities, which can help users to identify trends and patterns in their data.

Power BI is available as a cloud-based service, as well as in desktop and mobile versions, making it accessible from a variety of devices and platforms. Additionally, Power BI integrates with other Microsoft products such as Excel, SharePoint, and Teams, as well as with third-party tools and services.

DESIGN DETAILS

Layout: A well-designed dashboard should have a clear and consistent layout that is easy to navigate. Use a grid-based layout and align elements with a consistent margin to make your dashboard look organized and professional. Use headings, text boxes, and images to group and separate different sections of your dashboard.

Color: Use color wisely to draw attention to important information and make your dashboard visually appealing. Use a limited color palette and choose colors that are easy to read and complementary. Avoid using too many bright or contrasting colors, which can be distracting or hard on the eyes.

Typography: Use typography to create hierarchy and improve readability. Choose a font that is easy to read and consistent throughout your dashboard. Use font size, weight, and color to differentiate between headings, subheadings, and body text.

Visualization: Choose visualizations that effectively communicate your data and insights. Use charts, graphs, tables, and maps to present your data in a clear and compelling way. Use appropriate chart types based on the data you are presenting and avoid using too many different types of visualizations on the same dashboard.

Interactivity: Use interactivity to make your dashboard more engaging and informative. Use filters, slicers, and drill-down capabilities to allow users to explore the data in more detail. Use tooltips and annotations to provide additional context and information.

Performance: Ensure that your dashboard is fast and responsive. Use optimized data models, avoid complex calculations or queries, and limit the amount of data you load into your dashboard. Use caching and pre-aggregation to improve performance, especially for large datasets.

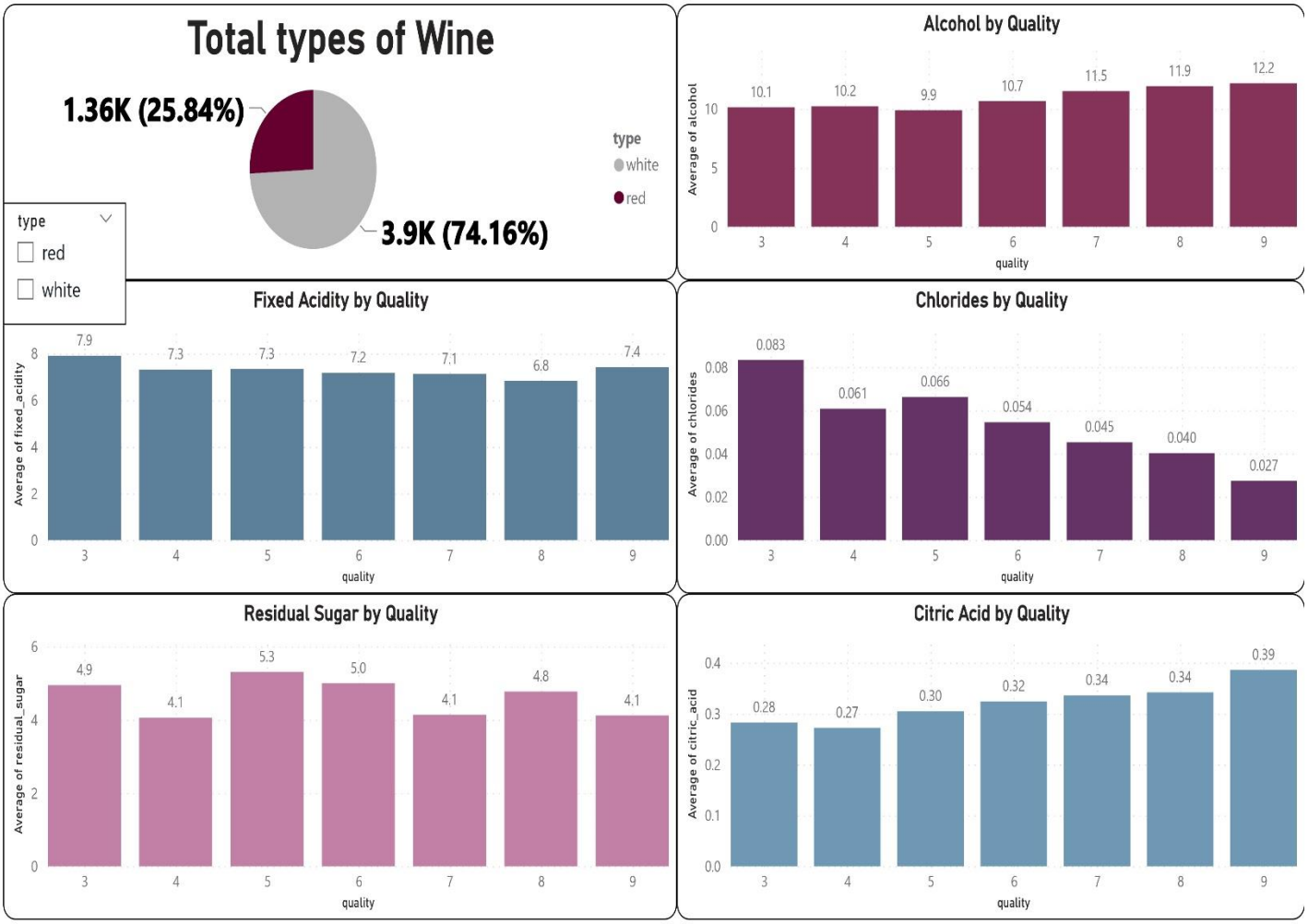
Accessibility: Make your dashboard accessible to all users, including those with disabilities. Use alt text for images, provide captions for videos, and use a high contrast mode for text and visuals. Use data tables for tabular data and ensure that your dashboard is keyboard-friendly.

OPTIMIZATION

Data Strategy and Performance.

- Removing duplicate records from the dataset.
- Handling null values.
- Performing feature encoding on the dataset.
- Developing model
- Building interactive filters to view data as required.
- Building Dashboard using Power BI.
- Building Predictor using Streamlit

Dashboard:


Total types of Wine
1.36K (25.84%)
3.9K (74.16%)
type
white
red
type
red
white
Alcohol by Quality
Average of alcohol
10.1
10.2
9.9
10.7
11.5
11.9
12.2
quality
Fixed Acidity by Quality
Average of fixed_acidity
7.9
7.3
7.3
7.2
7.1
6.8
7.4
quality
Chlorides by Quality
Average of chlorides
0.083
0.061
0.066
0.054
0.045
0.040
0.027
quality
Residual Sugar by Quality
Average of residual_sugar
4.9
4.1
5.3
5.0
4.1
4.8
4.1
quality
Citric Acid by Quality
Average of citric_acid
0.28
0.27
0.30
0.32
0.34
0.34
0.39
quality

Predictor:


Wine Quality Prediction Web App
By SHRUTI KHANDELWAL
Quantity of Fixed Acidity
1.00
1.00
20.00
Quantity of Volatile Acidity
0.00
0.00
2.00
Quantity of Citric Acid
0.00
0.00
5.00
Quantity of Residual Sugar
0.00
0.00
20.00
Quantity of Chlorides
0.00
0.00
2.00
Quantity of Free Sulfur Dioxide
0.00
0.00
100.00
Quantity of Total Sulfur Dioxide
0.00


Quantity of Free Sulfur Dioxide
0.00
0.00
100.00
Quantity of Total Sulfur Dioxide
0.00
0.00
200.00
Density
0.00
0.00
1.00
Value of pH
0.00
0.00
14.00
Quantity of Sulphates
0.00
0.00
10.00
Quantity of alcohol
0.00
0.00
20.00
Wine Quality Results
Wine Quality is Bad