Low Level Design (LLD)

Wine Data Analysis

****

## Revision Number - 1.2

**Last Date of Revision - 14/11/2022**

**SANJEEV KUMAR**

**Document Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Version** | **Description** | **Author** |
| 11/09/2022 | 1.0 | Introduction, Problem Statement | Sanjeev Kumar |
| 03/10/2022 | 1.1 | Dataset Information, Architecture Description | Sanjeev Kumar |
| 14/11/2022 | 1.2 | Final Revision | Sanjeev Kumar |

**2**

## Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Investment Analytics prediction analysis technique. It will explain the necessary steps which have to be followed before any analysis can begin. LLD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The LLD will be focusing on the below objectives:

* Problem Understanding.
* Data Acquisition.
* Data Pre-Processing and Exploratory Analysis
* Development of models
* Auditing accuracy and retrain if require
* Finalizing the model
* Dashboard report for important activities

**Scope**

The LLD documentation presents the detailed structure of the Investment analytics for each of its individual components. The goal of LLD is to give the internal logical design of the actual program code. Low-level design is created based on the high-level design. The LLD documentation contains the complete description of the model used along with the comparisons of the proposed model/library compared with a baseline(existing) model against a set of metrics.

**Project Introduction**

While the wine industry generally has been very proactive about dealing with climate change, from capturing fermentation carbon totrialling new varieties, this is ultimately a political problem. To find an effective solution for this problem you were asked to help in ETL process.

Do ETL : Extract-Transform-Load the dataset and find for me some information from this large data. This is form of data mining.

What all information can be achieved by mining this data, would bebrainstormed by the interns

Find key metrics and factors and show the meaningful relationships between attributes. Do your own research and come up with your findings.

**Constraints**

Our analysis is done based on a limited dataset provided for wines at sectors, regions basis.

**Risks**

Document specific risks that have been identified or that should be considered.

**Out of Scope**

Delineate specific activities, capabilities, and items that are out of scope for the project.

1. **Technical specifications**

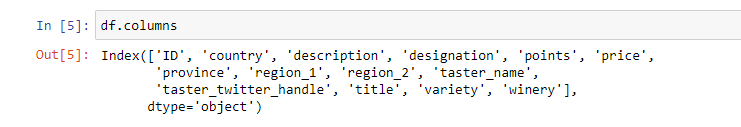
**Dataset**

The Dataset is taken from iNeuron’s provided dataset-



Figure 1: Wine Analytics Dataset

The dataset consists of 29665 rows data. There are 14 columns in the dataset which are mentioned below.



# Problem Statement

While the wine industry generally has been very proactive about dealing with climate change, from capturing fermentation carbon totrialling new varieties, this is ultimately a political problem. To find an effective solution for this problem you were asked to help in ETL process.

Do: Extract-Transform-Load the dataset and find for me some information from this large data. This is form of data mining.

Find key metrics and factors and show the meaningful relationships between attributes.

Do your own research and come up with your findings

# Architecture



Real World

Exploratory Data Analysis (EDA)

Modelling

Deployment

Data Cleaning

Data Pre- Processing

Raw Data Collection

Reporting

* 1. **Architecture Description**
     1. **Raw Data Collection-** The Dataset was taken from iNeuron provided Project Description Document.

https://drive.google.com/drive/folders/1FkmFVL8wlJmQWP1z52TD8PlhOJhitTyI?usp=s haring

### Data Pre-Processing

Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data to the model to train.

This Process includes-

* + - 1. Handling Null/Missing Values

### Data Cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

* + - 1. Remove duplicate or irrelevant observations
      2. Filter unwanted outliers
      3. Renaming required attributes

### Exploratory Data Analysis (EDA)

Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypothesis

and to check assumptions with the help of summary statistics and graphical representations.

### Reporting

Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in easy in report because your model will be used by many stakeholders who are not from technical background.

* + - 1. High Level Design Document (HLD)
      2. Low Level Design Document (LLD)
      3. Architecture
      4. Wireframe
      5. Detailed Project Report
      6. Power Point Presentation

### Modelling

Data Modelling is the process of analysing the data objects and their relationship to the other objects. It is used to analyse the data requirements that are required for the business processes. The data models are created for the data to be stored in a database. The Data Model's main focus is on what data is needed and how we have to organize data rather than what operations we have to perform.

### Deployment

