**High Level Design (HLD)**

**Wine Data Analytics**



**Revision Number - 1.2**

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

H HIGH LEVEL DESIGN (HLD)

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
|  |  |  |  |
| 12/09/2022 | 1.0 | Abstract, | Sanjeev Kumar |
|  |  | Introduction, |  |
|  |  | General Description |  |
|  |  |  |  |
| 02/10/2022 | 1.1 | Design Detail, KPI, | Sanjeev Kumar |
|  |  | Deployment |  |
|  |  |  |  |
| 14/11/2022 | 1.2 | Final Revision | Sanjeev Kumar |
|  |  |  |  |

**|** P a g e

Wine Data Analytics



**3 |** P a g e

HIGH LEVEL DESIGN (HLD)

**2**

**Contents**

Document Version Control…………………………………………………………………………2

Abstract……………………………………………………………………………………………….3

1. Introduction……………………………………………………………………………………….4

1.1 Why this High-Level Design Document?..............................................................4

1.2 Scope……………………………………………………………………………………4

1. General Description……………………………………………………………………………...5 2.1 Product Perspective & Problem Statement…………………………………………5

2.2 Tools Used……………………………………………………………………………...5

1. Design Detail……………………………………………………………………………………...6

3.1 Functional Architecture………………………………………………………………...6

3.2 Optimization………………………………………………………………………….….7

1. KPI……………………………………………………………………………………………….…8

4.1 KPIs (Key Performance Indicators) ……………………………………………..…...9

1. Deployment……………………………………………………………………………………......9

**** Wine Data Analytics



HIGH LEVEL DESIGN (HLD)

**1 Introduction**

**1.1 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 before coding and can be used as a reference manual for how the modules interact at a high level.

**The HLD will:**

* Present all of the design aspects and define them in detail
* Describe the user interface being implemented
* Describe the hardware and software interfaces
* Describe the performance requirements
* Include design features and the architecture of the project
* List and describe the non-functional attributes like:

-Security

-Reliability

-Maintainability

-Portability

-Reusability

-Application compatibility

-Resource utilization

-Serviceability

**1.2 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.

**2 General Description**

Wine Data Analytics **5 |** P a g e



HIGH LEVEL DESIGN (HLD)

**2.1 Product Perspective & 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 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.

**2.2 Tools used**

Business Intelligence tools and libraries works such as NumPy, Pandas, Matplotlib, MS-Excel, Power BI, Jupyter Notebook and Python Programming Language are used to build the whole framework.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

* Jupyter Notebook is used as IDE.
* Python is the Programming Language used.
* EDA is done using Numpy & Pandas.
* Visualizations were done using Matplotlib.
* Power BI is used for dashboard creation.

**3 Design Details**

**3.1 Functional Architecture**



**Assumptions**

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.

Our analysis assumes that all the data provided was true without any corruption and the features mentioned in the raw dataset are the only driving factors of a equilibrium investment.

HIGH LEVEL DESIGN (HLD)

**3.2 Optimization**

1. **Your data strategy drives performance**
   * Minimize the number of fields
   * Minimize the number of records
   * Optimize extracts to speed up future queries by materializing calculations, removing columns and the use of accelerated views
2. **Reduce the marks (data points) in your view**
   * Practice guided analytics. There’s no need to fit everything you plan to show in a single view. Compile related views and connect them with action filters to travel from overview to highly-granular views at the speed of thought.
   * Remove unneeded dimensions from the detail shelf.
   * Explore. Try displaying your data in different types of views.
3. **Limit your filters by number and type**
   * + Reduce the number of filters in use. Excessive filters on a view will create a more complex query, which takes longer to return results. Double-check your filters and remove any that aren’t necessary.
     + Use an include filter. Exclude filters load the entire domain of a dimension while including filters do not. An include filter runs much faster than an exclude filter, especially for dimensions with many members.
     + Use a continuous date filter. Continuous date filters (relative and range-ofdate filters) can take advantage of the indexing properties in your database and are faster than discrete data filters.
     + Use Boolean or numeric filters. Computers process integers and Booleans (t/f) much faster than strings.
     + Use parameters and action filters. These reduce the query load (and work across data sources).

**Performance**

Investment analytics determines the historic all the previous data and it should be as accurate as possible. So that it will not mislead to the future investor. Also, model retraining is very important to improve the performance.

**Security**

Since the investment care analysis consists of years data, the information should be secured.

**Reusability**

The code written and the components used should have the ability to be reused with no problems.

**Resource utilization**

When any task is performed, it will likely use all the processing power available until that function is finished.

**4 KPI**

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the Investment.

|  |
| --- |
|  |

As and when the system starts to capture the historical/periodic data for a Year, the dashboards will be included to display charts over time with progress on various indicators or factors

**4.1 KPIs (Key Performance Indicators)**

KPI (Key Performance Indicators) are those parameters that help us to analyze the data. Every data we want to use in Power BI has some parameters or KPI that drive the graph’s figures we want to see. We should monitor the KPIs of the data to analyze the changes, growth, downfall in the data. These KPIs become measurement points as well and help us in tracking the business. With the help of KPI visuals in Power BI, we can actually get the points that we want to hit to improve the business. We actually see, what target we have fixed and what actual figures we are getting on charts with the KPIs.

**5 Deployment**

Prioritizing data and analytics couldn’t come at a better time. Your company, no matter what size, is already collecting data and most likely analyzing just a portion of it to solve business problems, gain competitive advantages, and drive enterprise transformation. With the explosive growth of enterprise data, database technologies, and the high demand for analytical skills, today’s most effective IT organizations have shifted their focus to enabling self-service by deploying and operating Power BI at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

****