High Level Design (HLD) Analyzing Google Apps Store dataset

in terms of App downloads and Rating

Revision Number: 1.0

Last date of revision: 19/05/2024

Sairaj Honrao

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| **16thDec 2023** | 1.0 | First Version of Complete HLD | Sairaj Honrao |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Contents**

[Document Version Control 2](#_bookmark0)

[Abstract 3](#_bookmark1)

1. [Introduction 5](#_bookmark2)
   1. [Why this High-Level Design Document? 5](#_bookmark3)
   2. [Scope 5](#_bookmark4)
2. [General Description 6](#_bookmark5)
   1. [Product Perspective & Problem Statement 6](#_bookmark6)
   2. [Tools used 6](#_bookmark8)
3. [Design Details 7](#_bookmark9)
   1. [Functional Architecture 7](#_bookmark10)
   2. [Optimization 8](#_bookmark11)
4. [KPIs 9](#_bookmark12)
   1. [KPIs (Key Performance Indicators) 9](#_bookmark13)

# Abstract

This project leverages Microsoft Power BI to perform a comprehensive analysis of the Google Apps Store, aiming to uncover patterns, trends, and insights that can inform strategic decisions for app developers, marketers, and stakeholders. The analysis encompasses a variety of data points including app ratings, user reviews, download statistics, and category distributions.

Data was collected from the Google Apps Store using web scraping techniques and was subsequently cleaned and structured for analysis. Key performance indicators (KPIs) such as average rating, total downloads, and review sentiments were defined to provide a clear overview of the app market landscape.

The Power BI dashboards created in this project offer interactive visualizations that enable users to filter and drill down into specific data segments. These visualizations include:

Overall Market Trends: Displays general trends in app popularity and user engagement over time.

Category Analysis: Compares different app categories based on average ratings, total downloads, and other metrics.

Sentiment Analysis: Uses natural language processing (NLP) to analyze user reviews, highlighting common themes and user sentiments.

Geographical Insights: Shows download and rating distribution across different geographical regions.

The findings from this analysis reveal significant insights such as the correlation between app updates and user ratings, the impact of app size on download numbers, and the influence of user reviews on app popularity. These insights are valuable for app developers aiming to optimize their app performance and for marketers planning targeted campaigns.

# 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 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

## 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

## Product Perspective & Problem Statement

Housing prices are an important reflection of the economy, and housing price ranges are of great interest for both buyers and sellers. In this project, house prices will be predicted given explanatory variables that cover many aspects of residential houses.

The objective of the project is to perform data visualization techniques to understand the insight of the data. This project aims apply various Business Intelligence tools such as Power BI or Power BI to get a visual understanding of the data.

## Tools used

Business Intelligence tools and libraries works such as Excel, Power BI, Power BI are used to build the whole framework.





# Design Details

## Functional Architecture

|  |
| --- |
| Data Source (Excel File) |

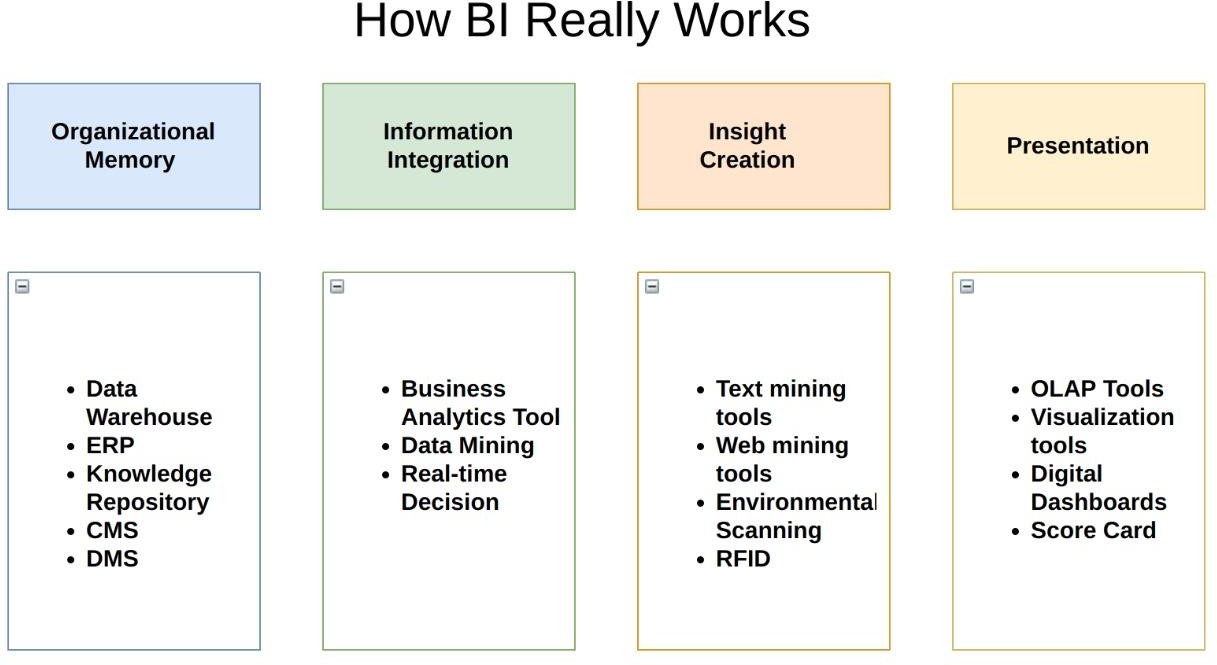
|  |
| --- |
| Data Preparation  (Power Query Editor)  - Import Data  - Data Cleaning - Transformation |

|  |
| --- |
| Data Model (Relationships,DAX  Measures  CalculatedColumns) |

|  |
| --- |
| Visualization  (Power BI Reports and Dashboards)  - Interactive Charts  - Slicers and Filters |

|  |
| --- |
| Distribution  (Power BI Service) - Publish Reports -Scheduled Refresh - Share with Stakeholders |

Figure 1: Functional Architecture of Business Intelligence



## Optimization

### 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.

### 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 include filters do not. An include filter runs much faster than an exclude filter, especially for dimensions with many members.

### Optimize and materialize your calculations

* + - Perform calculations in the database
    - Reduce the number of nested calculations.
    - Reduce the granularity of LOD or table calculations in the view. The more granular the calculation, the longer it takes.
      * LODs - Look at the number of unique dimension members in the calculation.
      * Table Calculations - the more marks in the view, the longer it will take to calculate.
    - [Where possible, use MIN or MAX instead of AVG](http://onlinehelp.tableau.com/current/pro/online/windows/en-us/help.htm#calculations_aggregation.html). AVG requires more processing than MIN or MAX. Often rows will be duplicated and display the same result with MIN, MAX, or AVG.
    - [Make groups with calculations](http://kb.tableau.com/articles/knowledgebase/creating-groups-using-calculated-fields). Like include filters, calculated groups load only named members of the domain, whereas Power BI’s group function loads the entire domain.
    - [Use Booleans or numeric calculations instead of string calculations](http://onlinehelp.tableau.com/current/pro/online/mac/en-us/help.htm#functions_functions_string.html). Computers can process integers and Booleans (t/f) much faster than strings. Boolean>Int>Float>Date>DateTime>String

# KPIs

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



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

## KPIs (Key Performance Indicators)

Key indicators displaying a summary of the Analyzing Google Apps Store dataset

in terms of App downloads and Rating with different metrics

* Orders by Age Group and Gender
* Orders by Gender
* Orders by Marital\_Status
* Total Customers
* Average Age
* Total Users
* Amount by Zone
* Product categories by Zone
* Orders by State and Zone
* Amount by State
* Amount by Zone

Power BI prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Power BI and Power BI Online leverage your existing technology investments and integrate into your IT infrastructure to provide a self-service, modern analytics platform for your users. With on-premises, cloud, and hosted options, there is a version of Power BI to match your requirements. Below is a comparison of the three types:

TYPE PROS CONS

### Power BI - On Premises

* Full control of hardware and software
* Infrastructure and data remain behind your firewall
* Need dedicated administrators to manage hardware and software
* Additional infrastructure needed to access off-network (mobile, external)

### Power BI - Public Cloud (IaaS)

* Full control of software on managed hardware
* Puts infrastructure in same place as data (for migration to cloud)
* Flexibility to spin up/down hardware as needed
* Need dedicated administrators to manage software
* Additional infrastructure needed to access off-network (mobile, external)

### Power BI Online (SaaS)

* Fully hosted solution (hardware, software upgrades)
* Fast to deploy
* Easy for external audience to access
* Single-site in multi-tenant environment
* Cubes are not supported
* No guest account access

Depending on your organizational roles and responsibilities, Power BI should be installed by a systems administrator and the designated Power BI Administrator in coordination with the appropriate IT roles. For Power BI Online, you will integrate with your existing technology and configure the site settings. The Data & Analytics Survey, completed by business teams, identifies and prioritizes data use cases, audience size, and users. You will use the information collected in both surveys to plan your deployment strategy, including sizing, installation, and configuration of your Power BI or integration and configuration of Power BI Online. In addition to installing Power BI or configuring Power BI Online, administrators will also need to plan for the client software installation of Power BI Prep Builder, Power BI Desktop, Power BI Mobile, and Power BI Bridge for Power BI Online where applicable.