High Level Design

Analysing Google App Store Dataset

Revision Number - 1.2

Last Date of Revision - 00/03/2023

Siddharth Bhat

Document Version Control

Date Issued	Version	Description	Author
00/03/2023	1.0	Abstract, Introduction, General Description	Siddharth Bhat
00/03/2023	1.1	Design Detail, KPI, Deployment	Siddharth Bhat
00/03/2023	1.2	Final Revision	Siddharth Bhat

Contents

ABSTRACT

This project aims to use the knowledge gained in Data Wrangling class on Goggle Play Store database. Google Play is a digital distribution service operated and developed by Google Inc. It serves as the official app store for the Android operating system, allowing users to browse and download applications developed with the Android SDK and published through Google. Google Play also serves as a digital media store, offering music, books, movies, and television programs books, movies, and television program

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

2.1 Product Perspective & Problem Statement

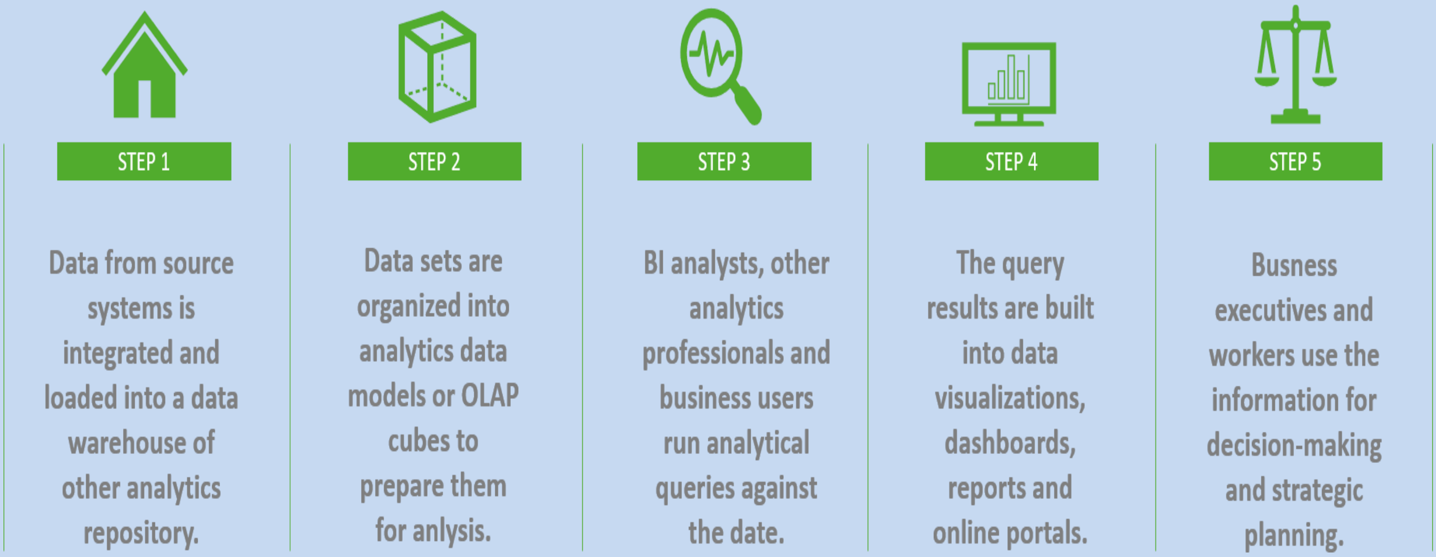
Technology is the increasing need nowadays and used everywhere. One of the features of Technology is android. Which we all use in our daily life. Android is a mobile operating system based on a modified version of the Linux kernel and other open source software, designed primarily for touchscreen mobile devices such as smartphones and tablets.

2.2 Tools used

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

3 Design Details

3.1 Functional Architecture



STEP 1

Data from source systems is integrated and loaded into a data warehouse of other analytics repository.

STEP 2

Data sets are organized into analytics data models or OLAP cubes to prepare them for anlysis.

STEP 3

BI analysts, other analytics professionals and business users run analytical queries against the date.

STEP 4

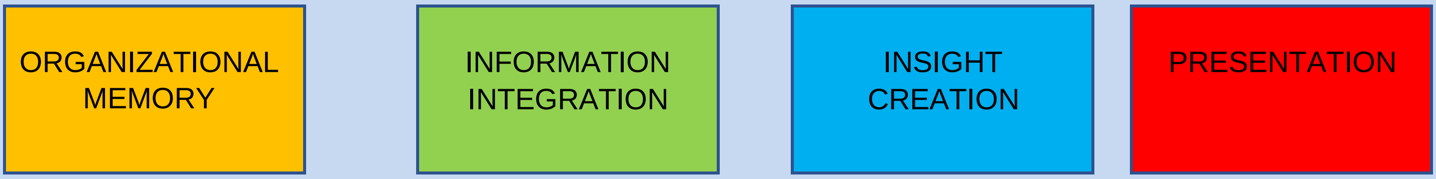
The query results are built into data visualizations, dashboards, reports and online portals.

STEP 5

Busness executives and workers use the information for decision-making and strategic planning.

Figure 1: Functional Architecture of Business Intelligence

How BI Works

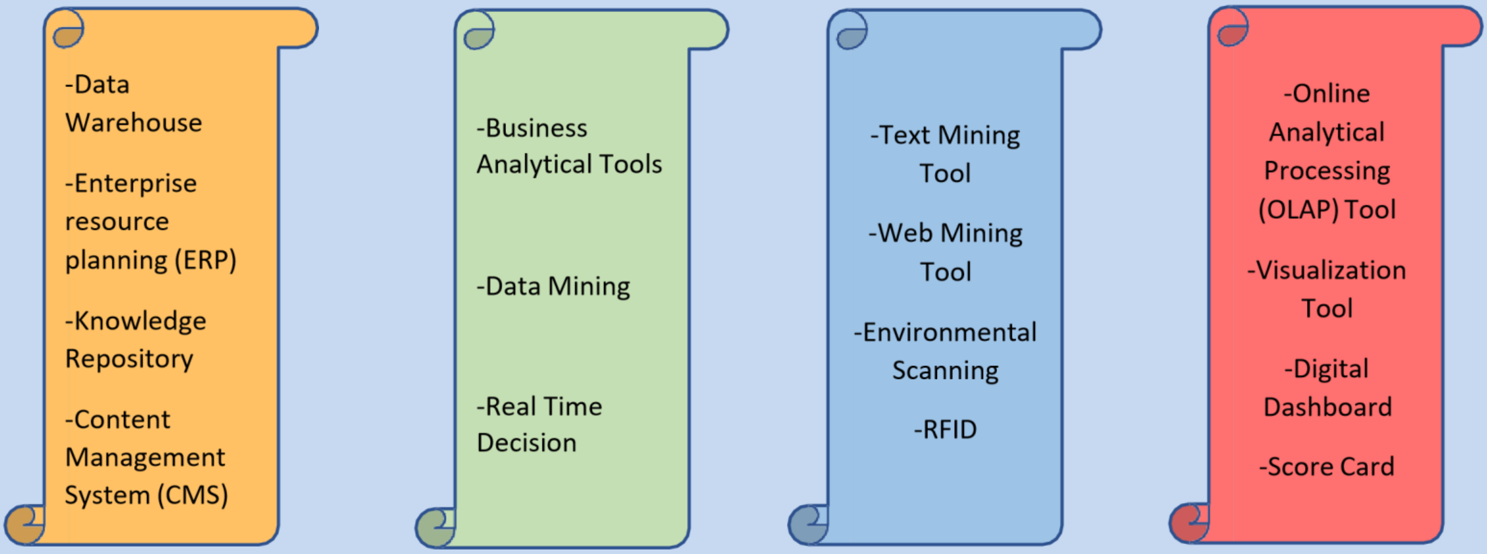


ORGANIZATIONAL MEMORY

INFORMATION INTEGRATION

INSIGHT CREATION

PRESENTATION



-Data Warehouse

-Enterprise resource planning (ERP)

-Knowledge Repository

-Content Management System (CMS)

-Business Analytical Tools

-Data Mining

-Real Time Decision

-Text Mining Tool

-Web Mining Tool

-Environmental Scanning

-RFID

-Online Analytical Processing (OLAP) Tool

-Visualization Tool

-Digital Dashboard

-Score Card

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-of date 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).

4. 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. 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. Like include filters, calculated groups load only named members of the domain, whereas Tableau’s group function loads the entire domain.

- Use Booleans or numeric calculations instead of string calculations. Computers can process integers and Booleans (t/f) much faster than strings. Boolean>Int>Float>Date>DateTime>String.

4 KPI

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

4.1 KPIs (Key Performance Indicators)

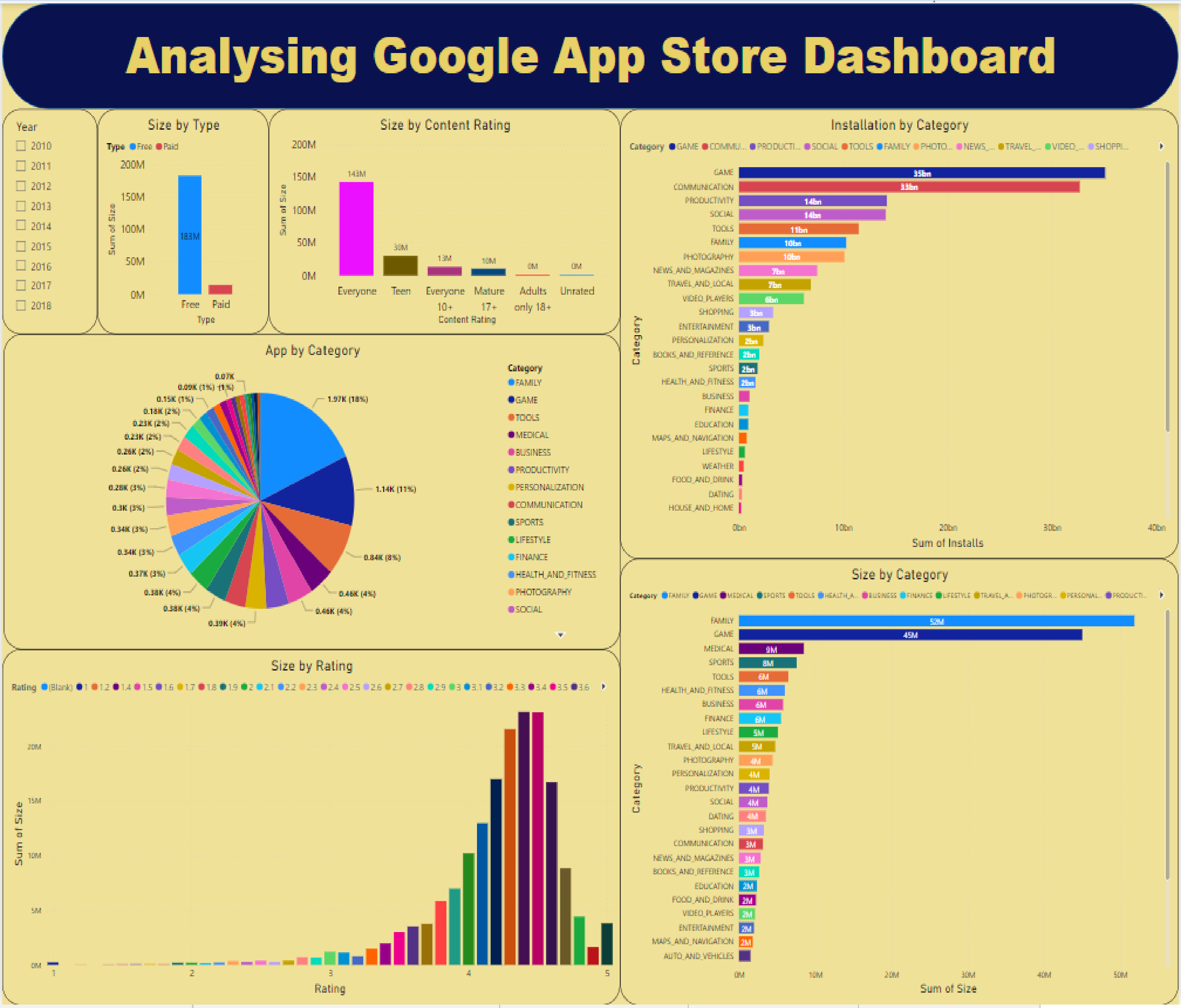
Key indicators displaying a summary of Sales Data and its relationships with different metrics.

1. Univariate Analysis of Categorical Features
2. Category wise App installation
3. Application Size by Category
4. Rating of Application
5. Reviews of Application

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 Analysing 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.

Power BI prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Power BI Desktop and Power BI Service leverage your existing technology investments and integrate them 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.


Analysing Google App Store Dashboard
Year
□ 2010
□ 2011
□ 2012
□ 2013
□ 2014
□ 2015
□ 2016
□ 2017
□ 2018
Size by Type
Type ● Free ● Paid
200M
150M
100M
50M
0M
Sum of Size
183M
Free
Paid
Type
Size by Content Rating
200M
150M
100M
50M
0M
Sum of Size
143M
30M
13M
10M
0M
0M
Everyone
Teen
Everyone 10+
Mature 17+
Adults only 18+
Unrated
Content Rating
Installation by Category
Category ● GAME ● COMMU... ● PRODUCTI... ● SOCIAL ● TOOLS ● FAMILY ● PHOTO... ● NEWS_... ● TRAVEL_... ● VIDEO_... ● SHOPPI...
GAME 35bn
COMMUNICATION 33bn
PRODUCTIVITY 14bn
SOCIAL 14bn
TOOLS 11bn
FAMILY 10bn
PHOTOGRAPHY 10bn
NEWS_AND_MAGAZINES 7bn
TRAVEL_AND_LOCAL 7bn
VIDEO_PLAYERS 6bn
SHOPPING 3bn
ENTERTAINMENT 3bn
PERSONALIZATION 2bn
BOOKS_AND_REFERENCE 2bn
SPORTS 2bn
HEALTH_AND_FITNESS 2bn
BUSINESS
FINANCE
EDUCATION
MAPS_AND_NAVIGATION
LIFESTYLE
WEATHER
FOOD_AND_DRINK
DATING
HOUSE_AND_HOME
Category
0bn
10bn
20bn
30bn
40bn
Sum of Installs
App by Category
0.07K
0.09K (1%) (1%)
0.15K (1%)
0.18K (2%)
0.23K (2%)
0.23K (2%)
0.26K (2%)
0.26K (2%)
0.28K (3%)
0.3K (3%)
0.34K (3%)
0.34K (3%)
0.37K (3%)
0.38K (4%)
0.38K (4%)
0.39K (4%)
0.46K (4%)
0.46K (4%)
0.84K (8%)
1.14K (11%)
1.97K (18%)
Category
● FAMILY
● GAME
● TOOLS
● MEDICAL
● BUSINESS
● PRODUCTIVITY
● PERSONALIZATION
● COMMUNICATION
● SPORTS
● LIFESTYLE
● FINANCE
● HEALTH_AND_FITNESS
● PHOTOGRAPHY
● SOCIAL
Size by Category
Category ● FAMILY ● GAME ● MEDICAL ● SPORTS ● TOOLS ● HEALTH_A... ● BUSINESS ● FINANCE ● LIFESTYLE ● TRAVEL_A... ● PHOTOGR... ● PERSONAL... ● PRODUCTI...
FAMILY 52M
GAME 45M
MEDICAL 9M
SPORTS 8M
TOOLS 6M
HEALTH_AND_FITNESS 6M
BUSINESS 6M
FINANCE 6M
LIFESTYLE 5M
TRAVEL_AND_LOCAL 5M
PHOTOGRAPHY 4M
PERSONALIZATION 4M
PRODUCTIVITY 4M
SOCIAL 4M
DATING 4M
SHOPPING 3M
COMMUNICATION 3M
NEWS_AND_MAGAZINES 3M
BOOKS_AND_REFERENCE 3M
EDUCATION 2M
FOOD_AND_DRINK 2M
VIDEO_PLAYERS 2M
ENTERTAINMENT 2M
MAPS_AND_NAVIGATION 2M
AUTO_AND_VEHICLES
Category
0M
10M
20M
30M
40M
50M
Sum of Size
Size by Rating
Rating ● (Blank) ● 1 ● 1.2 ● 1.4 ● 1.5 ● 1.6 ● 1.7 ● 1.8 ● 1.9 ● 2 ● 2.1 ● 2.2 ● 2.3 ● 2.4 ● 2.5 ● 2.6 ● 2.7 ● 2.8 ● 2.9 ● 3 ● 3.1 ● 3.2 ● 3.3 ● 3.4 ● 3.5 ● 3.6
20M
15M
10M
5M
0M
Sum of Size
1
2
3
4
5
Rating