

The background features a light beige color. In the top right corner, there is a large, light gray L-shaped block with a smaller, darker gray square nested within its corner. In the bottom left corner, there is a large, light gray L-shaped block with a smaller, darker gray square nested within its corner.

Low Level Design

ENTERTAINER DATA ANALYSIS
MANISHA RAJ A.

TABLE OF CONTENTS

1. Introduction	3
1.1 Why this Low-Level Document ?.....	3
1.2 Scope	3
2. Architecture	4
3. Architecture Description	4
3.1 Data Load & Cleaning	4
3.2 Initial Exploratory analysis	5
3.3 Data insights breakdown	6
3.4 Deployment	7
4. Unit Test Case	8

I. INTRODUCTION

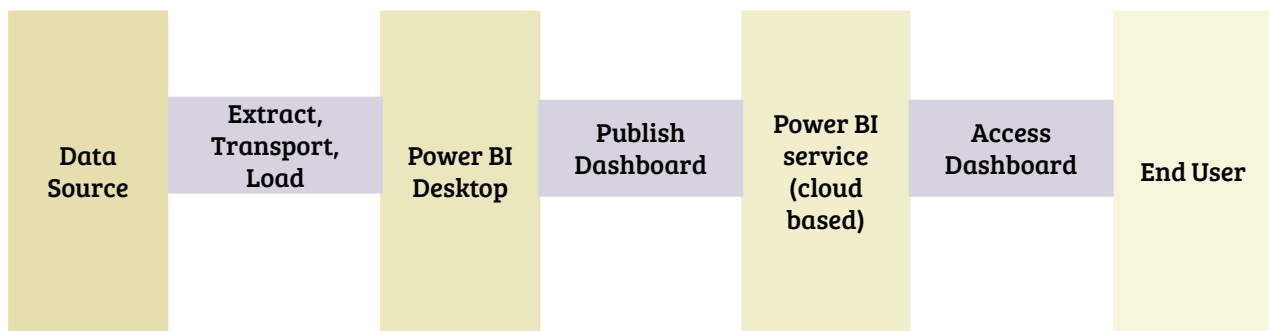
I.1 WHY THIS LOW-LEVEL DESIGN DOCUMENT?

The purpose of this Low-Level Design Document (LDD) is to provide detailed information about how the House Price Prediction dashboard is built. It includes diagrams that show the structure of the code, the relationships between different parts of the program, and specific details about how everything should work. This document helps programmers write the code by giving them a clear blueprint.

I.2 SCOPE

Low-Level Design (LLD) focuses on designing the individual components of the system. It involves creating detailed designs for data structures, software architecture, and performance algorithms. Initially, the overall data organization is defined during the requirement analysis phase, and then it is refined during the detailed design process.

2. ARCHITECTURE



3. ARCHITECTURE DESCRIPTION.

3.1. Data Loading and Cleaning

Components:

- Data Sources: The script retrieves datasets from Excel files hosted on GitHub, including [AdventureWorks_Database.xlsx](#) and [Budget.xlsx](#).
- Data Reading: Utilizes [pandas](#) to read multiple sheets from the Excel files into DataFrames.
- Data Cleaning:
 - Initial Display: Outputs the first few rows of each DataFrame to understand the structure.
 - Budget Data Cleaning: Removes unnecessary metadata and resets column headers.

- Merging DataFrames: Combines Sales, Customer, Product, and Territory data to create a comprehensive dataset.
- Missing Values Evaluation: Analyzes the percentage of missing values and visualizes this data using a bar plot.
- Further Cleaning: Drops columns with excessive missing values (over 40%) and any additional unuseful fields.

Tools:

- Pandas: For data manipulation and cleaning.
- Matplotlib & Seaborn: For visualizing missing data.

3.2 Initial Exploratory Analysis

Components:

- Demographic Insights:
 - Gender Distribution: Pie chart visualizing customer gender breakdown.
 - Marital Status Distribution: Pie chart for marital status.
 - Education Level: Count plot for customer education distribution.
- Sales Analysis:
 - Occupation vs. Sales Amount: Bar plot showing total sales by customer occupation.
 - Commute Distance vs. Sales Amount: Visualizes the relationship between commute distance and sales.
 - Category and Subcategory Distribution: Pie charts and count plots to analyze customer distribution by product categories.

Tools:

- Seaborn & Matplotlib: For creating various types of visualizations

3.3 Data Insight Breakdown

Components:

- Sales Performance Metrics:
 - Product Category and Subcategory Sales: Bar plots that illustrate total sales by category and subcategory.
 - Monthly Sales Trend: Time series plot to visualize sales trends over time.
 - Sales Performance by Year: Line plot showing yearly sales growth.
- Customer Behavior Analysis:
 - Sales Amount vs. Quantity Sold: Scatter plot to explore the relationship between order quantity and sales.
 - Average Sales by Education Level: Bar plot that highlights average sales per education category.
 - Income Group Analysis: Sales amount visualized by customer income brackets.

Tools:

- Pandas: For data aggregation and grouping.
- Seaborn & Matplotlib: For advanced visualizations.

3.4 Deployment

Deploying your Entertainer Data Analysis Dashboard in Power BI means making it accessible to users. Power BI offers several ways to share your dashboard:

1. **Power BI Service:** This is a cloud-based platform where you can publish, share, and collaborate on your dashboards. To deploy your dashboard here, you need to create a Power BI account, set up a workspace, and upload your dashboard file (PBIX) to that workspace. After uploading, you can adjust access permissions and share the dashboard with specific users or groups, allowing for collaboration.
2. **Power BI Desktop:** This tool is used for creating and designing your dashboard. If you want to deploy your dashboard using Power BI Desktop, you can save it as a PBIX file and share that file with others. Keep in mind that anyone who wants to view or interact with the dashboard will need to have Power BI Desktop installed on their computer.

This process ensures that dashboard is available and usable for the intended audience.

4. UNIT TEST CASES.

Test Case Description	Expectation Result
Test Data Loading	All datasets should load without errors, and their shapes should be correct.
Test Budget Data Cleaning	The cleaned Budget Data should not contain any metadata rows and should have a valid header.
Test Sales Data Column Dropping	The Sales Data should drop the specified columns without raising errors. The resulting DataFrame should have fewer columns.
Test Data Merging	The merged DataFrame should not contain any null values in the key columns and should maintain the expected number of rows.
Test Missing Value Analysis	It is Bar Graph which shows the Top 10 Entertainers from the Dataset
Test Dropping Columns with Missing Values	The columns dropped should exceed 40% missing values, and the remaining Data Frame should reflect this.
Test Gender Distribution Visualization	The pie chart should display without errors and correctly represent the gender distribution.
Test Sales Amount Grouping by Occupation	The sales amounts grouped by occupation should be correctly calculated and displayed in a bar plot.